

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
INSTITUTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA



Programa de Pós-Graduação  
em Química



Avaliação da termoestabilidade, atividade e resistência a ambientes ácidos  
de uma enzima de interesse biotecnológico via dinâmica molecular

**Sérgio Ruschi Bergamachi Silva**

Tese de Doutorado  
Natal/RN, agosto de 2018

SÉRGIO RUSCHI BERGAMACHI SILVA

Avaliação da termoestabilidade, atividade e resistência a ambientes ácidos de uma enzima de interesse biotecnológico via dinâmica molecular

Tese de doutorado apresentada ao Programa de Pós-graduação em Química, da Universidade Federal do Rio Grande do Norte, em cumprimento às exigências para obtenção do título de Doutor em Química.

Orientador: Prof. Dr. Davi Serradella Vieira

Natal, RN  
2018

Universidade Federal do Rio Grande do Norte - UFRN  
Sistema de Bibliotecas - SISBI  
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Silva, Sérgio Ruschi Bergamachi.

Avaliação da termoestabilidade, atividade e resistência a ambientes ácidos de uma enzima de interesse biotecnológico via dinâmica molecular / Sergio Ruschi Bergamachi Silva. - 2018. 135f.: il.

Tese (Doutorado)-Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Programa de Pós-Graduação em Química, Natal, 2018.

Orientador: Dr. Davi Serradella Vieira.

1. Xilanase - Tese. 2. Dinâmica molecular - Tese. 3. Mutantes - Tese. 4. Enzimas multidomínio - Tese. 5. Desenvolvimento biotecnológico - Tese. I. Vieira, Davi Serradella. II. Título.

RN/UF/BCZM

CDU 577.152.9

Sérgio Ruschi Bergamachi Silva

AVALIAÇÃO DA TERMOESTABILIDADE, ATIVIDADE E RESISTÊNCIA A  
AMBIENTES ÁCIDOS DE UMA ENZIMA DE INTERESSE BIOTECNOLÓGICO  
VIA DINÂMICA MOLECULAR

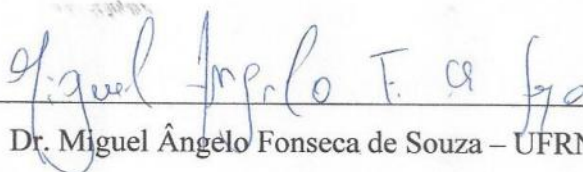
Tese apresentada ao Programa de Pós-graduação em Química da Universidade Federal do Rio Grande do Norte, em cumprimento às exigências para obtenção do título de Doutor em Química.

Aprovada em: 31 de agosto de 2018

Comissão Examinadora:



Dr. Davi Serradella Vieira – UFRN (orientador)



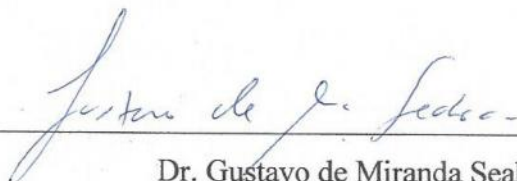
Dr. Miguel Ângelo Fonseca de Souza – UFRN



Dr. Euzébio Guimarães Barbosa – UFRN



Dr. Gutto Raffyson Silva de Freitas – IFRN



Dr. Gustavo de Miranda Seabra – UFPE

"A verdadeira viagem do descobrimento  
não consiste em procurar novas paisagens, mas em ter novos olhos."  
Marcel Proust

"A gente precisa continuar acreditando:  
que vale a pena ser honesto. Que vale a pena estudar. Que vale a pena trabalhar.  
Que é preciso construir: a vida, o futuro, o caráter, a família, as amizades e os  
amores."  
Lya Luft

"... the period of greatest gain in knowledge and experience is the most difficult  
period in one's life. If you go along in an easy way, with everything okay, you will feel  
that everything is just fine. Then, one day when you encounter problems you feel  
depressed and hopeless. Through a difficult period you can learn, you can develop  
inner strength, determination and courage to face the problem..."  
Dalai Lama

## AGRADECIMENTOS

Mesmo que não tenhamos a preocupação com o formalismo científico, regras da ABNT, ortografia perfeita e tantos outros padrões que nos é imposto no âmbito acadêmico, certamente a seção dos agradecimentos traz consigo uma característica ímpar. Ao mesmo tempo que é fácil e gostosa de escrever, justamente por estar livre de regras e revisões, é extremamente complexa visto que pessoas passam em nossa vida o tempo todo, e de forma singular deixam rastros diversos, as vezes muito pequenos, mas importantes. Esquecer alguém ou algo neste ponto geralmente é comum e, infelizmente, acabamos sendo ingratos com quem tanto nos ajudou. Perceber o crescimento ou a evolução que determinada pessoa te trouxe nem sempre é evidente e podemos demorar demais pra entender. Lembranças de qualquer espécie, seja de aprendizado, de conforto, compreensão, carinho, de paz ou de qualquer outro sentimento que nos faça lembrá-los. Sem mais delongas, seguem meus sinceros agradecimentos.

Ana Carolina. Achar as palavras certas para esse parágrafo tem seu grau máximo de dificuldade. É complicado expor em poucas linhas tudo que deveria e gostaria. Agradeço por partilharmos juntos os bons momentos, mas sou muito mais grato por me alavancar quando quase desistia. Sabes bem que nem entraria nesse doutorado se não tivesse sua compreensão e nem o concluiria sem seu apoio. Se existir uma palavra representativa esta seria: cumplicidade. Obrigado por ser parceira, forte, compreensiva e esposa. Sem dúvidas, por todos esses quatro anos de renúncias, você deveria ser co-autora desse material. Te amo!

Família. Só aqui poderia escrever várias páginas agradecendo por todos os menores momentos que fui ajudado, confortado e amado. Obrigado pela solidez que me foi dada durante toda minha criação. Meu pai, José Francisco e minha mãe, Dacira Pereira, os agradeço de coração. Obrigado também por toda torcida para que eu voltasse para Natal. Essa, sem dúvida, foi uma das nossas maiores vitórias. Amo vocês. Irmãs, Mila e Dany, vocês também têm um peso nessa tese. Obrigado por todo incentivo, compreensão e amor compartilhado. Que continuemos unidos e sendo bons filhos para nossos "velhinhos".

Agradeço também a duas figuras essenciais que tive a oportunidade de conhecer e trabalhar nesse período. O Prof. Zilvam e o Prof. Cláudio vocês foram

chefes sensacionais que possibilitaram o término mais suave dessa tese. Zilvam, sempre serei grato à você por todas as vezes que me liberou para assistir aulas, participar de reuniões de orientação ou simplesmente para estudar. Cláudio, também sou muito agradecido por todas as conversas, conhecimento compartilhado e toda compreensão durante minhas ausências do laboratório. Se não fosse isso, certamente o caminho teria sido bem mais duro.

Prof. Luiz Seixas e a Prof. Maria Bernardete. Serei eternamente grato por conhecê-los e por terem me mostrado a possibilidade de voltar pra casa. Isso transcende o período do doutoramento. A chegada à Natal foi essencial para conclusão desse material. Além deles, agradeço também ao Prof. Rodrigo Neves que tem sido compreensivo no tocante ao término do doutorado e compartilhado conversas produtivas após minha chegada ao ICe. Aliás, o ICe me apresentou três bons colegas de trabalho: Ana Cristina, Juliana Brandão e Joelson Dias. Agradeço pela companhia, pelo papo durante o dia e as guloseimas compartilhadas nesse último ano.

Aos amigos feitos na UFERSA, Atawalpa, Breno, Luiz Paulo, Mateus, Pedro Lucas, Rafael, Segundo e Vlagner, vocês desempenharam papel fundamental nessa jornada enquanto estivemos juntos. Tomaram conta do laboratório, montaram aulas enquanto estive ausente e me ensinaram muito durante as boas conversas. Os laços feitos foram muito importantes para revigorar minha vontade de avançar sempre. Obrigado caras.

Mesmo distante fisicamente, agradeço fortemente ao Carlos Fuzo que compartilhou seu conhecimento e material sobre as simulações de proteínas em diferentes pHs. Obrigado Carlos. Sua contribuição e paciência foram (e vão ser) fundamentais.

Agradeço os professores Euzébio Guimarães e Miguel Ângelo que compuseram minha banca de qualificação e deram contribuições fundamentais para melhoria do conteúdo dessa tese e aos professores Gustavo Seabra e Gutto Raffyson que vieram de longe para participar da construção desse material.

Não menos importante, agradeço de coração ao meu orientador Prof. Davi Serradella Vieira que tanto me ajudou nesse período. Apesar do início conturbado e de uma trajetória distante, a grande atenção e compreensão fizeram com que eu

não desistisse e pudesse desenvolver esse material. Obrigado por acreditar e por ter me aceitado como seu primeiro orientando de doutorado.

Por todos que não foram lembrados aqui na forma escrita mas que certamente estão dentro do meu coração, em especial aqueles que sempre perguntavam: "e o negócio lá termina quando?", agradeço pelos bons momentos vividos e por todo pensamento positivo compartilhado para o término dessa tese.



## RESUMO

A produção de etanol lignocelulósico através da rota enzimática tem ganhado espaço dentre os processos industriais a fim de substituir o tratamento ácido tradicional. As xilanases (E.C. 3.2.1.8) constituem uma classe de enzimas presentes no coquetel enzimático utilizado para esse fim e possuem grande interesse industrial/comercial devido sua alta versatilidade em diversos processos já conhecidos. Nesse contexto, algumas propriedades de interesse biotecnológico de uma xilanase da família GH11 foram avaliadas a partir de análises das simulações de dinâmica molecular (DM). Em um primeiro estudo, foram realizadas simulações por DM de uma xilanase produzida pela bactéria *Bacillus subtilis* (XynA\_WT) e um mutante quádruplo (Gln7His, Gly13Arg, Ser22Pro e Ser179Cys) que apresenta uma temperatura catalítica ótima 20 °C acima da XynA\_WT. Os resultados da DM sugerem possíveis estratégias para engenharia de xilanases da família GH11 para produzir enzimas termoestáveis. As mutações em regiões que apresentam pouca flexibilidade deveriam preservar a rigidez, mas a substituição delas pode ser escolhida para afetar favoravelmente outras propriedades, como a solvatação ou interações hidrofóbicas. Os dados não explicam apenas o efeito de termoestabilidade de uma xilanase de GH11 observado em experimentos anteriores de evolução direta, mas também fornece informações para o planejamento de outros mutantes GH11 termoestáveis através do desenho racional. Em um segundo estudo, o alvo foi uma quimera formada entre um domínio xilanolítico (XynA) e uma proteína receptora de xilose (XBP) que apresentou uma eficiência catalítica experimental quase 3,5 vezes maior que a mesma xilanase não-quimerizada. Os fatores responsáveis por essa discrepância foram entendidos através de simulações DM. Os resultados sugeriram a formação e estabilização de uma interface proteína-proteína entre os dois domínios na presença e na ausência de xilose no sítio ativo da XBP. Valores de energia potencial de interação (IPE) em função do tempo mostram que existe uma maior estabilização nas interações dessa interface para estrutura com xilose em comparação com a estrutura sem xilose. Parâmetros estruturais como a flexibilidade e volume do sítio ativo também foram avaliados. Em geral, os resultados sugerem que a quimera apresenta uma estrutura mais rígida em comparação com xilanase isolada e, em particular, a região do polegar, que controla a exposição do sítio ativo, demonstra uma redução significativa na flexibilidade. Por fim, simulações DM em diferentes pHs foram feitas com o intuito de entender a queda drástica na atividade catalítica da xilanase (isolada e na forma de quimera) quando presente em ambientes ácidos, e dessa forma, agregar informações para aumentar sua resistência nessas condições. Dados de mudanças do estado de protonação, distância abre-fecha do sítio ativo, raio de giro e superfície acessível ao solvente ajudaram a entender esse fenômeno e fornecem indicativos para futuras pesquisas nesse sentido. A tese reforça a capacidade descritiva, e posteriormente preditiva, das simulações DM no desenvolvimento biotecnológico de xilanases GH11.

Palavras-chave: xilanase; dinâmica molecular; mutantes: enzimas multidomínio; desenvolvimento biotecnológico

## ABSTRACT

The lignocellulosic ethanol production by enzymatic route has gained space among the industrial processes in order to replace the traditional acid treatment. Xylanases (E.C. 3.2.1.8) constitute a class of enzymes present in the enzymatic cocktail used for this purpose and have great industrial / commercial interest due to their high versatility in several known processes. In this context, some properties of biotechnological interest of a GH11 family xylanase were evaluated via molecular dynamics (MD) simulations. In a first study, we performed MD simulations of a xylanase produced by *Bacillus subtilis* (XynA\_WT) and a quadruple mutant (Gln7His, Gly13Arg, Ser22Pro and Ser179Cys) which display an higher optimum catalytic temperature (20°C) in relation to native. MD results suggest possible strategies for engineering GH11 xylanases to produce thermostable enzymes. Mutations in regions that exhibit reduced flexibility should preserve rigidity, but their substitution may be chosen to favorably affect other properties, such as solvation or hydrophobic interactions. The data do not only explain the thermostability effect of a GH11 xylanase observed in previous experiments of direct evolution, but also provide information for the planning of other thermostable GH11 mutants by rational design. In a second study, the target was a chimera formed between a xylanolytic domain (XynA) and a xylose binding protein (XBP) which had an experimental catalytic efficiency almost 3.5 times higher than the same non-chimerized xylanase. The factors responsible for this discrepancy were understood by MD simulations. The results suggested the formation and stabilization of a protein-protein interface between the two domains in the presence and absence of xylose in the active site of XBP. Interaction Potential Energy values (IPE) as a function of time show a greater stabilization in the interactions of this interface for xylose bound structure compared to the xylose free one. Structural parameters such as flexibility and volume of the active site were also evaluated. In general, the results suggest that the chimera displays greater rigidity in relation to free xylanase, and in particular the thumb region, which controls active site exposure, demonstrates a significant reduction in flexibility. Finally, MD simulations at different pHs were performed in order to understand the drastic decrease in the catalytic activity of this xylanase (native and chimeric form) in acidic environments, and, thus, to aggregate information to increase its resistance in such conditions. Data of side chain protonation states, open-close amplitude, gyration radius and solvent accessible surface are important analyzes to provide insights for elucidation of the mechanism of stabilization of this chimeras in acid environment. The thesis reinforces the descriptive and predictive capacities of the MD simulations in the biotechnological development of GH11 xylanases.

**Keywords:** xylanase; molecular dynamics; mutants; multidomain enzymes; biotechnological development

## LISTA DE FIGURAS

2.1	Representação de uma modulação alostérica a) positiva e b) negativa em sistema hipotético com ativador e inibidor alostérico.	29
3.1	Estrutura interna da parede celular das plantas evidenciando seus componentes: lignina, hemicelulose e celulose.	36
3.2	Estrutura representativa de uma parte da xilana. Cadeia principal formada por monômeros de D-xilopirranose (preto) e possíveis cadeias laterais de acetato (roxo), arabinofuranose (verde), ácido ferúlico (azul) e 4-O-metil-D-glicuronopiranosil (vermelho) com suas respectivas enzimas destacadas	38
3.3	Estrutura representativa de xilanases a) GH 10 e b) GH 11. O substrato está evidenciado próximo às cavidades catalíticas. $\alpha$ -hélices (roxo) e $\beta$ -folhas (amarelo) estão apresentadas em destaque.	40
3.4	Estrutura representativa de uma xilanase GH 11. Evidenciando regiões importantes e os resíduos catalíticos (vermelho).	41
4.1	Esquema representativo dos termos ligados, ou seja, a) estiramentos de ligação, b) deformação angular, c) variação diedral imprópria e d) variação diedral própria utilizado pelo campo de força GROMOS96.	60
4.2	Esquema representativo da replicação da caixa central de simulação em dois eixos devido a aplicação das PBCs	65
5.1	Representação 3D da enzima XynA-4x em dois ângulos distintos. Elementos da estrutura secundária foram definidos de acordo com as xilanases <i>T. reesei</i> I e II	81
5.2	Flutuações quadráticas médias (RMSF) por resíduo da XynA_WT (quadrado preto) e XynA-4x (círculo branco) calculadas em 328 K e 348 K para os átomos da cadeia principal das (A) posições 5–9; (B) posições 10–16; (C) posições 17–27 e (D) posições 175–185.	84
6.1	Representação da quimera XynA–XBP. a) A sequência da quimera XynA–XBP271. Regiões da XBP estão apresentadas em azul e as da XynA em vermelho. O número entre colchetes indica o número do aminoácido nas proteínas isoladas. b) Representação tridimensional do modelo final obtido via simulações DM da quimera XynA–XBP271.	97
6.2	Raiz do desvio médio quadrático (RMSD) ao longo do tempo simulado da quimera XynA–XBP271 e dos seus domínios na a) ausência e b) presença de xilose no sítio de ligação da XBP.	99

6.3	Estruturas iniciais e finais provenientes das simulações DM para a quimera XynA–XBP271 a) com xilose e b) sem xilose no sítio de ligação da XBP.	100
6.4	Energia potencial de interação (IPE) interdomínios (xilanase e XBP) na quimera XynA–XBP271 com xilose (linha preta) e sem xilose (linha cinza). Dados obtidos via simulações DM ao longo de 120 ns.	101
6.5	Resultado do <i>Alanine Scanning</i> computacional mostrando os resíduos mais importantes (acima do limiar de 1,0 kcal/mol indicado pela linha tracejada vermelha) para estabilização da interface proteína-proteína entre os domínios da XBP (destaque em azul) e Xilanase (destaque em vermelho) obtidos após simulações DM. $\Delta\Delta G_{\text{ligação}}$ para mutações com alanina em resíduos da quimera XynA-XBP271 na presença (barras pretas) e ausência (barras cinza) de xilose.	102
6.6	Raiz da flutuação quadrática média (RMSF) por resíduo da quimera XynA-XBP271 com xilose (linha preta), sem xilose (linha vermelha) e da xilanase isolada (linha verde).	103
6.7	<i>Essential dynamics</i> das quimeras representadas por setas em uma estrutura 3D representativa a partir das simulações DM. As setas indicam a direção dos movimentos locais das cadeias proteicas no domínio de xilanase (vermelha) nas quimeras a) XynA–XBP271_cX e b) XynA–XBP271_sX.	104
6.8	Quantidade de aminoácidos interfaciais nas quimeras XynA-XBP209_cX (com xilose, vermelho contínuo), XynA-XBP209_sX (sem xilose, vermelho pontilhado), XynA-XBP271_cX (com xilose, preto contínuo) e XynA-XBP271_sX (sem xilose, preto pontilhado).	108
6.9	Diferença na quantidade de aminoácidos interfaciais nas quimeras com e sem xilose. Quimera XynA-XBP209 (vermelho) e XynA-XBP271 (preto).	109
6.10	Representação de pontes salinas inter e intradomínios na quimera XynA-XBP271 com o domínio xilanolítico representado em azul (com a região do polegar destacada) e o domínio XBP, em vermelho. Aminoácidos Asp267 e Asp390 representados em forma de tubos vermelhos. Aminoácido Arg344 em forma de tubos verdes.	110
6.11	Energia potencial de interação da ponte salina Asp390-Arg344 (em kcal/mol) na xilanase isolada (linha verde) e nas quimeras XynA-XBP271 com e sem xilose, XynA-XBP271_cX (linha preta) e XynA-XBP271_sX (linha vermelha), respectivamente.	111
6.12	Energia potencial de interação da ponte salina Asp267-Arg344 (em kcal/mol) nas quimeras XynA-XBP271 com e sem xilose, XynA-XBP271_cX (linha preta) e XynA-XBP271_sX (linha vermelha), respectivamente.	111

<b>6.13</b>	Energia potencial de interação entre os aminoácidos da interface que se distanciam no máximo 8 Å do outro domínio. Dados obtidos via simulação DM para os resíduos de XBP.	114
<b>7.1</b>	Esquema representativo do funcionamento da simulação CpHMD utilizada nesta tese para determinar os estados de protonação das cadeias laterais na quimera XynA-XBP271 e da XynA_WT.	122
<b>7.2</b>	RMSD calculado tomando como referência os C $\alpha$ da quimera XynA-XBP271 em pH 4,5 (verde) e 6,5 (azul) e para XynA_WT em pH 4,5 (preto) e 6,5 (vermelho) ao longo de 50 ns de simulação.	123
<b>7.3</b>	Raio de giro calculado em longo de 50 ns para a quimera XynA-XBP271 em pH 4,5 (verde) e 6,5 (azul) e para XynA_WT em pH 4,5 (preto) e 6,5 (vermelho).	124
<b>7.4</b>	Representações da enzima XynA_WT em diferentes pHs. A) Estruturas tridimensionais finais após 50 ns de simulação DM sobrepostas em dois ângulos distintos (visão lateral à esquerda e visão frontal do sítio ativo à direita). Os dois resíduos catalíticos (E78 e E172) estão evidenciados em tubos vermelhos. Estrutura em pH 4,5 em azul claro e pH 6,5 em azul escuro. B) Superfície de potencial eletrostático para XynA_WT em pH 6,5 (à esquerda) e pH 4,5 (à direita). Visão lateral do sítio ativo. Variação do vermelho (cargas mais negativas) para o azul (cargas mais positivas).	125
<b>7.5</b>	Distância Abre-Fecha (em Angstroms) em longo de 50 ns para a quimera XynA-XBP271 em pH 4,5 (verde) e 6,5 (azul) e para XynA_WT em pH 4,5 (preto) e 6,5 (vermelho). A medida avaliada entre o polegar e a região dos dedos (Fig. 3.4).	126
<b>7.6</b>	Representação dos aminoácidos que tiveram o estado de protonação das cadeias laterais alterados quando comparadas as simulações em pH 6,5 e 4,5 da quimera XynA_XBP271. Em destaque os aminoácidos do domínio da xilanase (XynA, em azul). Histidinas representadas em tubos pretos, glutamatos em tubos amarelos e aspartatos em tubos verdes. Os aminoácidos Asp354, Glu349, Glu443, His420 e His427 correspondem respectivamente à Asp83, Glu78, Glu172, His149 e His156 na simulação da XynA_WT.	127
<b>7.7</b>	Mecanismo de reação representativo para formação do complexo enzima-substrato. Substrato xilana representada como grupo R. Etapas envolvendo reagentes (RT), estados de transição (TS1 e TS2), intermediário (INT) e produto (P) até esse ponto da reação.	129

**7.8** Representação da visão frontal do sítio ativo da XynA (da quimera XynA-XBP271) destacando a blindagem dos glutamatos catalíticos para os pHs 6,5 (à esquerda) e 4,5 (à direita). Estrutura secundária em azul, 130 resíduos catalíticos, Glu349 e Glu443, em tubos vermelhos e uma superfície representativa da blindagem existente (em verde).

## LISTA DE TABELAS

- 5.1** Energia Potencial de Interação dos resíduos 7 e 13 com a vizinhança (raio de 6,0 Å) obtidos através das simulações DM em 328K e 348K para XynA\_WT e XynA-4x. 86
- 6.1** Parâmetros cinéticos experimentais de eficiência catalítica e efeito de ativação da xilose para quimeras XynA-XBP209, XynA-XBP262 e XynA-XBP271 106
- 7.1** Fração de tempo que as cadeias laterais dos resíduos estiveram na sua forma protonada. Valores descritos para a quimera XynA\_XBP271 e XynA\_WT em pH 6,5 e 4,5. 128

## ABREVIATURAS, ACRÔNIMOS E SIGLAS EMPREGADAS

<b>AMBER</b>	Campo de força ou pacote de programas de simulação; do inglês, <i>Assisted Model Building with Energy Refinement</i>
<b>ASAS</b>	Área Superficial Acessível ao Solvente
<b>CHARMM</b>	Campo de força ou pacote de programas de simulação; do inglês, <i>Chemistry at HARvard Macromolecular Mechanics</i>
<b>CNPEM</b>	Centro Nacional de Pesquisa em Energia e Materiais
<b>CpHMD</b>	Dinâmica molecular com pH constante; do inglês, <i>Constant-pH Molecular Dynamics</i>
<b>CTBE</b>	Laboratório Nacional de Ciência e Tecnologia do Bioetanol
<b>DFT</b>	Teoria do Funcional da Densidade; do inglês, <i>Density Functional Theory</i>
<b>DM</b>	Dinâmica Molecular
<b>DNA</b>	Ácido desoxirribonucleico
<b>EAX</b>	Efeito de Ativação da Xilose
<b>ECat</b>	Eficiência Catalítica
<b>ECat<sub>relativa</sub></b>	Eficiência Catalítica relativa
<b>IPE</b>	Energia Potencial de Interação; do inglês, <i>Interaction Potential Energy</i>
<b>EUA</b>	Estados Unidos da América
<b>GH11</b>	Glicosil Hidrolase, família 11
<b>Gln7His</b>	Mutação de glutamina por histidina na posição 7
<b>Gly13Arg</b>	Mutação de glicina por arginina na posição 13
<b>GROMACS</b>	Software para cálculo de trajetórias da dinâmica molecular; do inglês, <i>GROningen MACHine for Chemical Simulations</i>
<b>HB</b>	Ligações de hidrogênio; do inglês, <i>hydrogen bonds</i>
<b>IUBMB</b>	União Internacional de Bioquímica e Biologia Molecular; do inglês, <i>International Union of Biochemistry and Molecular Biology</i>
<b>LAMMPS</b>	Software para cálculo de trajetórias da dinâmica molecular; do inglês <i>Large-scale Atomic/Molecular Massively Parallel Simulator</i>
<b>LINCS</b>	Algoritmo de restrição; do inglês, <i>Linear Constraint Solver</i>



<b>MCTI</b>	Ministério da Ciência, Tecnologia e Inovação
<b>MWC</b>	Modelo Monod-Wyman-Changeux
<b>NAMD</b>	Software para cálculo de trajetórias da dinâmica molecular; do inglês, <i>Nanoscale Molecular Dynamics</i>
<b>NIPE</b>	Núcleo Interdisciplinar de Planejamento Energético
<b>NVT</b>	Ensemble com número de partículas (N), volume (V) e temperatura (T) constantes
<b>PBC</b>	Condições Periódicas de Contorno; do inglês, <i>Periodic Boundary Conditions</i>
<b>PDB</b>	Banco de dados de proteínas; do inglês, <i>Protein Data Bank</i>
<b>PME</b>	Método de cálculo do potencial eletrostático da família do somatório de Ewald; do inglês, <i>Particle Mesh Ewald</i>
<b>PROPKA</b>	Script para cálculo de pKa de estruturas biológicas
<b>RF</b>	Campo de reação; do inglês, <i>Reaction Field</i>
<b>RMSD</b>	Raiz do Desvio Quadrático Médio; do inglês, <i>Root Mean Square Deviation</i>
<b>RMSF</b>	Raiz da flutuação quadrática média; do inglês, <i>Root Mean Square Fluctuation</i>
<b>RNA</b>	Ácido ribonucleico
<b>SEP</b>	Superfície de Energia Potencial
<b>Ser179Cys</b>	Mutação de serina por cisteína na posição 179
<b>Ser22Pro</b>	Mutação de serina por prolina na posição 22
<b>SPC</b>	Modelo de água de ponto de carga simples; do inglês <i>Simple Point Charge</i>
<b>SPME</b>	Método de cálculo do potencial eletrostático da família do somatório de Ewald; do inglês, <i>Smooth Particle Mesh Ewald</i>
<b>T<sub>OPT</sub></b>	Temperatura catalítica ótima
<b>vdW</b>	van der Waals
<b>VMD</b>	Software de visualização de estruturas; do inglês, <i>Visual Molecular Dynamics</i>
<b>XBP</b>	Proteína receptora de xilose; do inglês, <i>Xilose Binding Protein</i>
<b>XynA_WT</b>	Xilanase GH11 de <i>Bacillus subtilis</i> nativa, mesofílica
<b>XynA- 4x</b>	Mutante quádruplo de xilanase termofílica GH11
<b>XynA-XBP209</b>	Quimera XynA-XBP fundidas na posição 209 da XBP
<b>XynA-XBP209_cX</b>	Quimera XynA-XBP fundidas na posição 209 da XBP e com xilose ligada

<b>XynA-XBP209_sX</b>	Quimera XynA-XBP fundidas na posição 209 da XBP e sem xilose ligada
<b>XynA-XBP262</b>	Quimera XynA-XBP fundidas na posição 262 da XBP
<b>XynA-XBP271</b>	Quimera XynA-XBP fundidas na posição 271 da XBP
<b>XynA-XBP271_cX</b>	Quimera XynA-XBP fundidas na posição 271 da XBP e com xilose ligada
<b>XynA-XBP271_sX</b>	Quimera XynA-XBP fundidas na posição 271 da XBP e sem xilose ligada

## PUBLICAÇÕES DURANTE O PERÍODO DE DOUTORAMENTO

- i. Ribeiro, L. F.; Tullman, J.; Nicholes, N.; **Silva, S. R. B.**; Vieira, D. S.; Ostermeier, M.; Ward, R. J. *A xylose-stimulated xylanase–xylose binding protein chimera created by random nonhomologous recombination*. *Biotechnology for Biofuels* (2016), 9, 1, 119.  
QUALIS – A1\*
- ii. Morais, C. L. M.; **Silva, S. R. B.**; Vieira, D. S.; Lima, K. M. G. *Integrating a Smartphone and Molecular Modeling for Determining the Binding Constant and Stoichiometry Ratio of the Iron(II)–Phenanthroline Complex: An Activity for Analytical and Physical Chemistry Laboratories*. *Journal of Chemical Education* (2016) 93,10, 1760–1765.  
QUALIS – A2\*
- iii. Firme, C. L.; Monteiro, N. K. V.; **Silva, S. R. B.** *QTAIM and NCI analysis of intermolecular interactions in steroid ligands binding a cytochrome P450 enzyme–Beyond the most obvious interactions*. *Computational and Theoretical Chemistry* (2017) 1111, 40-49.  
QUALIS – B2\*
- iv. **Silva, S. R. B.**; Pinheiro, M. P.; Fuzo, C. A.; Silva, S. R.; Ferreira, T. L.; Lourenzoni, M. R.; Nonato, M. C.; Vieira, D. S.; Ward, R. J. *The role of local residue environmental changes in thermostable mutants of the GH11 xylanase from *Bacillus subtilis**. *International Journal of Biological Macromolecules* (2017) 97, 574-584.  
QUALIS – A2\*

\* Dados referentes à Classificação de Periódicos no Quadriênio de 2013-2016 e área de avaliação Química.

## SUMÁRIO

<b>1.</b>	<b>Etanol Lignocelulósico: contexto histórico, situação atual e perspectivas futuras</b>	<b>20</b>
<b>2.</b>	<b>Proteínas – Estrutura, Função e Efeito Alostérico</b>	<b>24</b>
<u>2.1</u>	<u>Estrutura e Função</u>	<u>25</u>
<u>2.2</u>	<u>Estabilidade Proteica</u>	<u>27</u>
<u>2.3</u>	<u>Efeito Alostérico</u>	<u>28</u>
<b>3.</b>	<b>Xilanases – Histórico e Aplicações</b>	<b>35</b>
<u>3.1</u>	<u>Substrato</u>	<u>36</u>
<u>3.2</u>	<u>Classificação e características estruturais</u>	<u>38</u>
<u>3.3</u>	<u>Mutações e xilanases multidomínio</u>	<u>40</u>
<u>3.4</u>	<u>Aplicações</u>	<u>42</u>
<u>3.4.1</u>	<i>Produção do bioetanol</i>	<i>42</i>
<b>4.</b>	<b>Fundamentos da Dinâmica Molecular</b>	<b>49</b>
<u>4.1</u>	<u>Conceitos de simulação molecular</u>	<u>52</u>
<u>4.1.1</u>	<i>Equações de movimento</i>	<i>53</i>
<u>4.1.2</u>	<i>Modelos de solvatação</i>	<i>56</i>
<u>4.1.3</u>	<i>Algoritmos de restrição</i>	<i>57</i>
<u>4.2</u>	<u>Campos de Força</u>	<u>57</u>
<u>4.2.1</u>	<i>Estiramento de ligação</i>	<i>59</i>
<u>4.2.2</u>	<i>Deformação angular</i>	<i>59</i>
<u>4.2.3</u>	<i>Variações diedrais</i>	<i>60</i>
<u>4.2.4</u>	<i>Interações não-ligadas de van der Waals</i>	<i>62</i>
<u>4.2.5</u>	<i>Interações eletrostáticas não-ligadas</i>	<i>62</i>
<u>4.2.6</u>	<i>Raio de corte e condições periódicas de contorno</i>	<i>63</i>
<u>4.3</u>	<u>Ferramentas de Análise</u>	<u>66</u>
<u>4.3.1</u>	<i>Raiz do desvio quadrático médio (do inglês, Root Mean Square Deviation, RMSD)</i>	<i>67</i>
<u>4.3.2</u>	<i>Raiz da flutuação quadrática média (do inglês, Root Mean Square Fluctuation, RMSF)</i>	<i>67</i>
<u>4.3.3</u>	<i>Avaliação da distância</i>	<i>68</i>

4.3.4	<i>Energia de interação</i>	68
4.3.5	<i>Quantidade de Resíduos Interfaciais</i>	69
4.3.6	<i>Alanine Scanning Computacional</i>	70
4.3.7	<i>Área Superficial Acessível ao Solvente (ASAS)</i>	71
<b>5.</b>	<b>O efeito de mutações locais em uma xilanase GH11 termoestável: uma análise por dinâmica molecular</b>	<b>77</b>
<u>5.1</u>	<u>Detalhes Computacionais</u>	<u>80</u>
<u>5.2</u>	<u>Resultados e Discussão</u>	<u>82</u>
5.2.1	<i>Mutação Q7H</i>	82
5.2.2	<i>Mutação G13R</i>	83
5.2.3	<i>Mutação S22P</i>	87
5.2.4	<i>Mutação S179C</i>	88
<u>5.3</u>	<u>Conclusão</u>	<u>90</u>
<b>6.</b>	<b>Análise via dinâmica molecular da estabilidade e movimentação de uma quimera xilanase-XBP</b>	<b>95</b>
<u>6.1</u>	<u>Detalhes Computacionais</u>	<u>97</u>
<u>6.2</u>	<u>Quimeras XynA-XBP271_cX e XynA-XBP271_sX</u>	<u>99</u>
6.2.1	<i>Resultados e Discussão</i>	99
<u>6.3</u>	<u>Comparação entre diferentes quimeras</u>	<u>105</u>
6.3.1	<i>Resultados e Discussão</i>	107
<u>6.4</u>	<u>Conclusão</u>	<u>114</u>
<b>7.</b>	<b>Dependência do pH na atividade xilanolítica de uma enzima multidomínio</b>	<b>119</b>
<u>7.1</u>	<u>Conceito das Simulações <i>Constant-pH MD</i> (CpHMD)</u>	<u>120</u>
<u>7.2</u>	<u>Metodologia e parâmetros de simulação</u>	<u>121</u>
<u>7.3</u>	<u>Resultados e Discussão</u>	<u>123</u>
<u>7.4</u>	<u>Conclusão</u>	<u>131</u>
	<b>Conclusões Gerais</b>	<b>134</b>

---

## Capítulo 1

### **Etanol Lignocelulósico: contexto histórico, situação atual e perspectivas futuras**

A preocupação com o esgotamento do petróleo aliado às questões ambientais e sociais evidentes têm elevado o interesse em energias renováveis, e o etanol de segunda geração, bioetanol ou etanol lignocelulósico, é um dos principais candidatos à substituição dos combustíveis fósseis. O etanol de segunda geração foi pela primeira vez produzido por Henri Braconnot em 1819, quando descobriu que era possível hidrolisar a celulose em açúcares pelo tratamento com ácido sulfúrico e posteriormente fermentar através de micro-organismos. Um curto início de comercialização desse combustível foi feito na Alemanha em 1898 e até houve a construção de instalações nos Estados Unidos em 1910 para essa produção mas todas acabaram fracassando por dificuldades econômicas.

No século seguinte, a demanda por fontes de energias diferentes do petróleo cresceu demasiadamente com o apelo ambiental por produtos mais ecologicamente corretos, sendo necessário o fomento à pesquisa por combustíveis “verdes”. Assim, a produção de etanol cresceu e teve incentivo de vários governos em diversas partes do mundo, inclusive no Brasil com o programa Proálcool na década de 1970. Mais recentemente, outro grande exemplo disso foi o discurso de estado feito por George W. Bush em janeiro de 2006 propondo a expansão definitiva do etanol celulósico e em 2007, a proposta de produção de 130 bilhões de litros de (bio)etanol até 2017. Simultaneamente, a demanda por alimentos e, conseqüentemente, a necessidade de mais terras para produzi-los também aumentou, gerando um certo conflito de interesses que até hoje dificulta a competitividade do álcool mundialmente.

Vários tipos de plantações vem sendo utilizadas para essa produção de etanol e variam de acordo com a localização do país e as condições climáticas existentes. O mosto produzido tem diversas procedências como o milho (China e EUA), trigo (China e Europa), cana de açúcar (Brasil), beterraba (Europa).

Atualmente, no Brasil, a produção de etanol a partir da sacarose de cana-de-açúcar está restrita à exploração de cerca de 35% do valor energético total da matéria-prima. Esse rendimento poderia ser significativamente aumentado por uma tecnologia que mobilizasse os outros 65% de energia, que está presente na forma de biomassa lignocelulósica. A produção mundial de biomassa de celulose foi estimada em cerca de 1,5 trilhões de toneladas/ano, caracterizando-a como uma fonte inesgotável de matéria-prima para a produção de biocombustíveis minimizando a competição entre a produção alimentar e a de combustíveis.

A lignocelulose, formada por lignina, hemicelulose e celulose, é o maior componente estrutural dos organismos vegetais (biomassa). Para a fermentação do substrato e posterior obtenção do álcool, a hemicelulose e a celulose precisam ser separadas da lignina e hidrolisadas em monômeros de pentoses e hexoses para a posterior fermentação por micro-organismos. Inicialmente é feito um pré-tratamento para reduzir as porções de biomassa e aumentar a superfície de contato para as posteriores reações (químicas e enzimáticas), além de um tratamento (geralmente) ácido para fragmentar as fibras existentes. Depois disso, acontece a sacarificação do material (hemi)celulósico, ou seja, a hidrólise é feita por enzimas (também conhecidas como celulasas e hemicelulasas) ou por agentes químicos (ácido ou base) para reduzir a celulose em açúcares simples. Nesse segundo processo que reside o entrave tecnológico impedindo a produção em larga escala. Agentes inorgânicos fortes podem ser usados para esse fim, mas acabam gerando baixos rendimentos e produtos indesejáveis durante o processo, tornando o bioetanol inviável econômica e ecologicamente quando produzido por essa rota.

A saída proposta passa através de rotas enzimáticas mais suaves (ecologicamente) para a hidrólise da (hemi)celulose. O problema dessa rota de produção é o alto custo associado à produção em larga escala do coquetel enzimático.

O processo enzimático foi inspirado em enzimas presentes no organismo de ruminantes que fazem esse processo digestivo adaptadas à indústria.

As enzimas responsáveis por esse processo são chamadas de celulasas e hemicelulasas. As enzimas celulasas atuam diretamente na hidrólise da celulose, enquanto que as hemicelulasas, são consideradas enzimas acessórias desse processo, pois atuam no substrato hemicelulósico, como as xilanases, xilosidases, glucuronidases, arabinofuranosidases e outras.

Foi necessária a criação de um método para que as enzimas pudessem atuar juntas em todas as moléculas específicas dentro do reator industrial, devido à complexidade dessa matéria-prima, principalmente da hemicelulase. Um coquetel enzimático para degradar o material lignocelulósico foi proposto, ou seja, várias celulasas e hemicelulasas devem agir sinergicamente na degradação da celulose e hemicelulose.

No intuito de aumentar a produtividade dessas enzimas e baratear os custos de produção, estudos têm sido desenvolvidos em torno delas para o aumento da sua reatividade e termoestabilidade. O aumento da termoestabilidade de enzimas tem sido desenvolvido há tempos e com grande sucesso através da mutação-sítio dirigida. O mesmo acontece para reatividade de reações enzimáticas. Para tanto, a engenharia de proteína é necessária. Através do estudo prévio do comportamento da enzima em determinado meio, os pesquisadores têm encontrado padrões de movimentação e interação os quais podem ser otimizados através de mutações em aminoácidos específicos. O conhecimento para executar essas substituições experimentalmente é bem consolidado no meio dos bioquímicos principalmente, tornando essa parceria cada vez mais viável e produtiva.

A ferramenta utilizada para entendimento a nível molecular do comportamento e dos processos enzimáticos relacionados às proteínas é a química computacional. Estudos teóricos sobre a dinâmica, estabilidade e reatividade dessas enzimas são de suma importância dentro do contexto da engenharia de proteínas. O entendimento, a nível molecular, dos fatores que conduzem a uma maior estabilidade, o controle local da rigidez/flexibilidade e o mapeamento das interações intermoleculares proteína-substrato, proteína-proteína e proteína-solvente, nos possibilita prever e desenhar novas enzimas mais adaptadas a uma determinada aplicação biotecnológica. As pesquisas atualmente estão usando também métodos quânticos e híbridos para observar interações entre o substrato e as enzimas. As simulações quânticas e híbridas também já estão sendo executadas e a tendência é



que continuam a expandir-se com o avanço dos *softwares* e *hardwares* disponíveis no mercado. A engenharia de proteínas tem usado esse recurso como ferramenta para o desenvolvimento de enzimas celulasas e hemicelulasas mais aprimoradas para o uso industrial.

Dessa forma, a produção de bioetanol deverá ser barateada através do uso de enzimas mais termoestáveis, que se manterão no reator por mais tempo em temperaturas mais elevadas, ou através de reações mais eficazes que trarão maior rendimento gastando menos energia.

A possibilidade de aumentar a produção de etanol a partir do bagaço da cana-de-açúcar e de outros materiais celulósicos num país de grande riqueza agrícola tem impulsionado os estudos sobre as enzimas envolvidas nesse processo. Os estudos computacionais são importantes e também constituem as metodologias empregadas na engenharia de proteínas de modo a direcionar as modificações genéticas de forma mais racional e eficaz. Os estudos computacionais sobre as enzimas envolvidas no processo ainda são escassos e podem complementar e enriquecer os resultados experimentais.

No Brasil, além das universidades, grande parcela dos resultados experimentais na linha de etanol de 2º geração é produzida pelo Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE). A concepção do CTBE aconteceu em 2005 quando o Ministério da Ciência, Tecnologia e Inovação (MCTI), por meio de uma pesquisa exploratória estratégica encomendada ao Núcleo Interdisciplinar de Planejamento Energético (NIPE), decidiu investir em pesquisas para a expansão da produção nacional de etanol. A grande missão do CTBE é a transferência dos processos desenvolvidos em escala laboratorial para o setor produtivo. O CTBE foi inaugurado em 2010 e está integrado ao Centro Nacional de Pesquisa em Energia e Materiais (CNPEM). Desde então, as linhas de pesquisas que contemplam o etanol de 2º geração têm recebido grande atenção e caráter estratégico por parte das agências de fomentos, que incentivam desde a pesquisa básica, com lançamento de editais gerais (bionergia) ou específicos, até a criação de empresas de bases tecnológicas.

---

## Capítulo 2

### Proteínas – Estrutura, Função e Efeito Alostérico

Os seres vivos têm seus processos biológicos mais importantes realizados em nível molecular. O bom funcionamento desses organismos passa por uma cinética e termodinâmica adequada das reações (bio)químicas existentes, ou seja, devem durar um tempo aceitável e demandar uma energia razoável. Além disso, fatores estruturais e dinâmicos das estruturas microscópicas devem estar devidamente definidos a fim de que os demais processos biológicos ocorram espontaneamente. Para que isso ocorra, as proteínas possuem um papel fundamental, como a regulação do balanço osmótico de membranas, transporte de agentes químicos, manutenção da estrutura do corpo celular, dentre outros. Contudo, provavelmente a maior contribuição dessas macromoléculas seja na catálise de reações bioquímicas vitais ao metabolismo (ALMÉN et al., 2009; COPELAND, 2004; GOODSELL; OLSON, 2000; NELSON; COX, 2004; PERLAND; FREDRIKSSON, 2017; PRESLAND; DALE, 2000).

As proteínas que possuem a característica de catalisar reações são agrupadas como enzimas. Excetuando algumas moléculas de RNA que catalisam reações específicas, todas as enzimas são proteínas (NELSON; COX, 2004; SOARES; LINS, 1995). Assim como os demais catalisadores, esses agentes diminuem a energia de ativação de reações biológicas as quais demandariam mais energia ou levariam muito tempo para ocorrerem.

As enzimas em geral são altamente específicas em suas reações. Os diferentes aminoácidos que as constituem podem gerar infinitas possibilidades estruturais dependendo de como estão dispostos espacialmente. Essa característica

aliada à flexibilidade das suas cadeias forma um ambiente propício e único para interações específicas com ligantes ou entre proteínas.

A especificidade das proteínas é proveniente de interações, na maioria das vezes, não covalentes, entre determinados grupos proteicos e/ou não-proteicos. Essas interações entre diferentes regiões da proteína pode inclusive modular movimentos, e consequentemente, a atividade de enzimas. Esse efeito derivado da comunicação entre diferentes domínios/regiões da mesma enzima após interação com um determinado ligante é conhecido como alosterismo ou efeito alostérico e apresenta-se por diferentes mecanismos em aplicações variadas.

Nesse capítulo serão abordados conceitos fundamentais sobre a bioquímica das proteínas. Inicialmente, na seção 2.1, sobre a sua formação, constituição e classificação de estruturas. Em seguida, sua estabilidade sob ação de fatores externos na seção 2.2 e por fim, os conceitos e aplicações do efeito alostérico na seção 2.3.

## 2.1 Estrutura e Função

As proteínas têm a sua estrutura e função diretamente correlacionadas. A função proteica depende fundamentalmente da estrutura na qual a proteína se enovela e como se dispõe sob solvatação. Por isso, saber a constituição de cada proteína é essencial para se ter mais detalhes a nível molecular de parâmetros como tamanho, solubilidade e peso molecular, por exemplo.

Essas macromoléculas são formadas a partir de aminoácidos individuais, também conhecidos como resíduos de aminoácido (ou ainda, somente resíduos), que se ligam covalentemente através de ligações peptídicas, e por isso as proteínas são consideradas polímeros naturais. Os monômeros, neste caso os aminoácidos, começaram a ser descobertos em 1806 e hoje já se conhece 20 deles, se considerarmos somente os naturalmente encontrados (NELSON; COX, 2004). Os resíduos possuem características em comum, a saber, um carbono assimétrico (carbono  $\alpha$ ) ligado a um hidrogênio, um grupo carboxila e um grupo amino, além de um grupo R que os torna diferentes.

Essa diferença entre os resíduos torna as proteínas altamente peculiares em relação às suas funções e estrutura. Eles podem ser classificados de acordo com

seu grupo R que podem apresentar cargas positivas, negativas ou nulas, podem ter caráter hidrofílico ou hidrofóbico, podem ser alifáticos ou aromáticos e diferirem também entre tamanhos. A união de vários resíduos de mesma característica pode formar cavidades hidrofóbicas ou regiões de alta densidade eletrônica que direcionam reações altamente específicas. Além disso, alguns aminoácidos também podem comportar-se como ácidos ou bases (natureza anfotérica) se adequando à reações e substratos dependendo do pH e solvente do meio no qual estão.

As proteínas podem ser descritas em função da organização linear desses diferentes aminoácidos. A sequência desses monômeros da forma com que acontecem as ligações peptídicas e sem importar-se com a disposição espacial deles, é conhecida como estrutura primária. Ela é especialmente importante na comparação de proteínas, já que proteínas da mesma classe compartilham boa parte da sequência de aminoácidos, mantendo principalmente determinadas regiões cruciais para sua função. São muito utilizadas na bioinformática como informações de entradas para as buscas nos bancos de dados disponíveis (BOCK; GOUGH, 2001; KELLEY; STERNBERG, 2009; WHISSTOCK; LESK, 2003).

No início do século XX demonstrou-se que os aminoácidos não estavam linearmente esticados e que possuíam portanto, configurações flexíveis como as dobras e torsões (NELSON; COX, 2004). Logo, pôde-se classificar arranjos estáveis energética e estruturalmente que se repetem em diversas proteínas de maneira recorrente, no caso as alfa-hélices ( $\alpha$ -hélices) e conformações beta (ou  $\beta$ -folhas). Essas estruturas (conhecidas como secundárias) são estabilizadas por meio das ligações de hidrogênio entre carbonilas e grupos amino, sejam dentro de um mesmo resíduo ou entre resíduos próximos.

Além da interação por ligações de hidrogênio, os aminoácidos podem interagir através de suas cadeias laterais, as quais apresentam grande diversidade de polaridade, tamanho e carga, resultando em um aumento na flexibilidade da cadeia polimérica de forma mais acentuada. Logo, a descrição completa da disposição tridimensional de todos os aminoácidos da proteína é chamada de estrutura terciária (NELSON; COX, 2004; VOET; VOET; PRATT, 1999). Apesar de bem descritas pelas três estruturas anteriormente apresentadas, certas proteínas ainda precisam de uma quarta subdivisão para que sejam completamente representadas. Algumas delas possuem mais de uma cadeia peptídica com estrutura primária, secundária e

terciária diferentes entre si, mas que se agregam para formar uma única estrutura. Esse arranjo espacial de subunidades é conhecido como estrutura quaternária (NELSON; COX, 2004; VOET; VOET; PRATT, 1999).

## 2.2 Estabilidade Proteica

As proteínas possuem uma estabilidade limitada ao ambiente que estão. As estruturas, divididas em diferentes níveis hierárquicos, são estabilizadas por interações dentro da própria cadeia polipeptídica, e/ou através da interação com cofatores e solvente. Essas interações vão desde as ligações covalentes de alta estabilidade na estrutura primária até as fracas e curtas interações hidrofóbicas nas quaternárias.

A modificação do meio que circunda determinada proteína pode torná-la inativa pela modificação das interações intra e intermoleculares, processo denominado desnaturação. Esse processo pode ser entendido como a perda da atividade biológica naturalmente exercida pela proteína devido sua menor organização estrutural (MURPHY, 2001). Assim sendo, fatores externos podem quebrar as interações anteriormente existentes tornando a cadeia de aminoácidos estruturalmente menos organizada e, por consequência, inespecífica quanto à sua função (SHOICHET et al., 1995).

Fatores como pH (modificam a carga das cadeias laterais possibilitando a repulsão entre grupos de aminoácidos), temperatura (sua elevação enfraquece as interações intermoleculares responsáveis pela estrutura secundária, terciária e quaternária) e solventes (modificam a constante dielétrica do meio afetando o enovelamento da cadeia hidrofóbica) são os mais importantes para que a desnaturação ocorra (BRANDTS; OLIVEIRA; WESTORT, 1970; GRIEBENOW; KLIVANOV, 1996; KUMAR; NUSSINOV, 2001; TANFORD, 1968; WALLQVIST; COVELL; THIRUMALAI, 1998; YANG; HONIG, 1993). Por esses motivos, o ambiente no qual a maioria das proteínas estão (com exceção de enzimas altamente peculiares, conhecidas como extremófilas (ADAMS; PERLER; KELLY, 1995; DEMIRJIAN; MORÍS-VARAS; CASSIDY, 2001; ELLEUCHE et al., 2014; HOUGH; DANSON, 1999)) apresentam valores de pH próximos da neutralidade e temperatura ambiente, além de estarem repletos de água. As proteínas que

conseguem ser ativas em uma faixa mais ampla desses fatores, são aplicadas comercialmente e/ou têm grande importância biológica.

A água é um dos poucos solventes que possui propriedades físicas e químicas para estabilizar o meio biológico. Por esse motivo é o maior constituinte na maioria dos seres vivos e está presente nos fluidos corporais, órgãos, células e afins. Sua molécula possui um volume muito pequeno e uma massa baixa quando comparada à lipídios e proteínas. Esses fatores a deixam com uma viscosidade e densidade ideais para os processos bioquímicos como o transporte de sais por canais iônicos por exemplo. Além disso, apresenta uma baixa tendência de ionização e alta polaridade que propiciam estabilidade em ambientes altamente carregados e favorecem as interações, principalmente através de ligações de hidrogênio, com as cadeias laterais dos aminoácidos (JEFFREY; SAENGER, 1991; NELSON; COX, 2004).

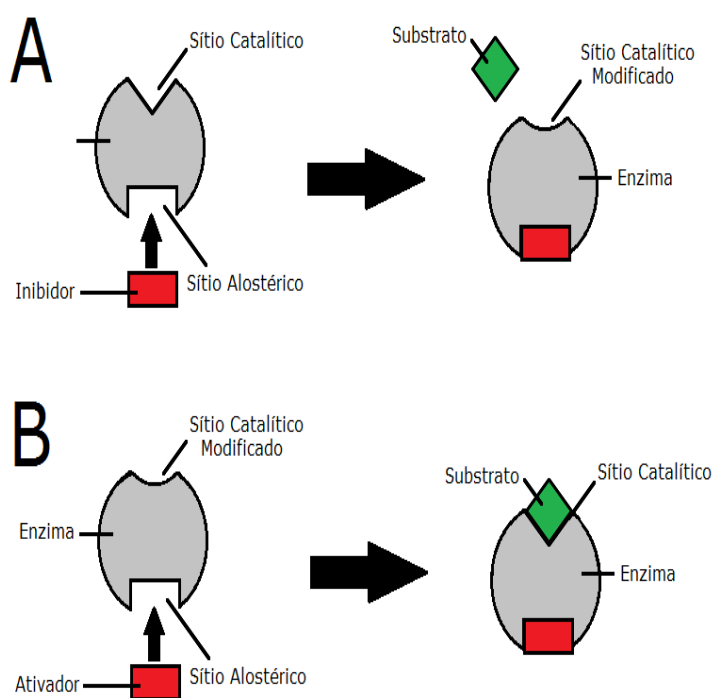
### 2.3 - Efeito Alostérico

Algumas proteínas têm sua movimentação ou atividade biológica modulada através de fatores externos. Esse efeito causado em uma região ou em toda proteína devido a interação com outra molécula, seja ela pequena ou outra proteína, é conhecido como alosterismo (efeito alostérico) (BU; CALLAWAY, 2011; HILSER; WRABL; MOTLAGH, 2012). A visão geral de que um ligante pode afetar a conformação da proteína alvo, não é nova; ela foi proposta por Wyman no final da década de 40 usando a hemoglobina como modelo (WYMAN; ALLEN, 1951; WYMAN, 1948). Desde então, esse fenômeno também tem sido demonstrado para outras proteínas (FREDERICK et al., 2007; HANSON et al., 2007). Embora o efeito tenha uma essência estrutural, a nova conformação pode apresentar uma atividade melhor ou pior que a original. Dentro das células, vários eventos que conduzem à mudanças conformacionais de proteínas, induzidas por ligantes, podem ocorrer. Enzimas se ligam a seus substratos, receptores se ligam a seus hormônios cognatos, anticorpos a seus antígenos e etc.

Essa regulação alostérica pode ser positiva ou negativa e varia de acordo com a proteína, modulador e função que desempenha. A modulação positiva ocorre quando um modulador alostérico (ativador), designação dada à molécula (ou

macromolécula) que promove esse efeito, liga-se ao sítio alostérico de determinada proteína de modo a aumentar ou possibilitar sua atividade catalítica. Já a modulação negativa faz exatamente o oposto ao diminuir ou anular a atividade correspondente. Em ambos os casos, a proteína é dita como alostérica (CHANGEUX; CHRISTOPOULOS, 2016; SCHWAB et al., 2017; SRINIVASAN et al., 2014).

Como exemplificado na Fig. 2.1, o sítio alostérico é diferente daquele onde ocorre a reação, neste caso, o sítio catalítico, da enzima. Entre os dois sítios há uma distância considerável na qual ocorrem mudanças, muitas vezes, com alta movimentação das cadeias internas, e assim, o efeito é gerado e propagado.



**Figura 2.1** – Representação de uma modulação alostérica a) positiva e b) negativa em sistema hipotético com ativador e inibidor alostérico.

Vários modelos foram publicados a fim de explicar os efeitos encontrados nas proteínas alostéricas, dentre os quais se destacam o modelo concertado e o sequencial. O modelo criado por Monod, Wyman e Changeux, também conhecido como modelo MWC, a enzima é dividida em subunidades que apresentam sempre a mesma conformação, seja relaxada (favorável a entrada do substrato) ou tensa (desfavorável a entrada do substrato). Esses dois tipos de conformação estão presentes no sistema em uma razão definida por um equilíbrio termodinâmico e na

medida em que há a entrada de ligantes nas estruturas relaxadas, o equilíbrio é deslocado favorecendo uma das conformações. Isso acontece até total consumo do substrato e reestabelecimento da razão original entre as estruturas relaxada e tensa (MONOD; WYMAN; CHANGEUX, 1965).

Já no modelo sequencial, criado por Koshland, Némethy e Filmer, a enzima possui subunidades com conformações diferentes que são alteradas após entrada do ligante. Inicialmente o sistema possui estruturas tensionadas que são convertidas em relaxadas após interação aleatória com o substrato. A partir de então, conformações das subunidades da enzimas que estão/são conectadas mudam sua estrutura tornando-se favoráveis à entrada de novas moléculas ligantes de forma sequenciada (KOSHLAND; NÉMETHY; FILMER, 1966). Vários outros modelos existem para descrever esse efeito que, obviamente, não se limitam aos exemplos dado acima (JAFFE, 2005;; LEVINE; WEINSTEIN, 2015;; MOTLAGH et al., 2014). De um modo geral, as proteínas alostéricas possuem dois sítios de ligação que induzem uma mudança na sua estrutura buscando um mínimo de energia do complexo enzima-ligante que resulta em modulações positivas ou negativas.

Essa interação entre modulador e sítio alostérico pode ter naturezas diferentes dependendo do substrato e da reação que a enzima catalisa. Caso a molécula reguladora seja o próprio substrato que se liga ao sítio ortoestérico e alostérico, trata-se de um modulador homotrópico, enquanto que os heterotrópicos são moléculas diferentes daquelas que se ligam ao sítio ativo da enzima. Diversos casos de moduladores homotrópicos e heterotrópicos já foram reportados na literatura e possuem diversas aplicações desde o metabolismo celular em nosso corpo, enzimas alostéricas com aplicação industrial até no design de drogas que se utilizam desse mecanismo (GUNTAS; OSTERMEIER, 2004; MONOD; CHANGEUX; JACOB, 1963; VILLAVERDE, 2003; WANG et al., 2013; ZIEGLER, 1985).



## Referências - Capítulo 2

- ADAMS, M. W. W.; PERLER, F. B.; KELLY, R. M. Extremozymes: Expanding the Limits of Biocatalysis. **Nature Biotechnology**, v. 13, n. 7, p. 662–668, 1 jul. 1995.
- ALMÉN, M. et al. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. **BMC Biology**, v. 7, n. 1, p. 50, 2009.
- BOCK, J. R.; GOUGH, D. A. Predicting protein-protein interactions from primary structure. **Bioinformatics**, v. 17, n. 5, p. 455–460, 1 maio 2001.
- BRANDTS, J. F.; OLIVEIRA, R. J.; WESTORT, C. Thermodynamics of protein denaturation. Effect of pressure on the denaturation on ribonuclease A. **Biochemistry**, v. 9, n. 4, p. 1038–1047, fev. 1970.
- BU, Z.; CALLAWAY, D. J. E. Proteins MOVE! Protein dynamics and long-range allostery in cell signaling. v.83, p. 163–221, 2011.
- CHANGEUX, J.-P.; CHRISTOPOULOS, A. Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation. **Cell**, v. 166, n. 5, p. 1084–1102, ago. 2016.
- COPELAND, R. A. **Enzymes: a practical introduction to structure, mechanism, and data analysis**. [s.l.] John Wiley & Sons, 2004.
- DEMIRJIAN, D. C.; MORÍS-VARAS, F.; CASSIDY, C. S. Enzymes from extremophiles. **Current Opinion in Chemical Biology**, v. 5, n. 2, p. 144–151, abr. 2001.
- ELLEUCHE, S. et al. Extremozymes—biocatalysts with unique properties from extremophilic microorganisms. **Current Opinion in Biotechnology**, v. 29, p. 116–123, out. 2014.
- FREDERICK, K. K.; MARLOW, M. S.; VALENTINE, K. G.; WAND, A. J. Conformational entropy in molecular recognition by proteins. **Nature**, vol. 448, p. 325–330, 2007.
- GOODSELL, D. S.; OLSON, A. J. Structural Symmetry and Protein Function. **Annual Review of Biophysics and Biomolecular Structure**, v. 29, n. 1, p. 105–153, jun. 2000.
- GRIEBENOW, K.; KLIBANOV, A. M. On Protein Denaturation in Aqueous–Organic Mixtures but Not in Pure Organic Solvents. **Journal of the American Chemical Society**, v. 118, n. 47, p. 11695–11700, jan. 1996.
- GUNTAS, G.; OSTERMEIER, M. Creation of an Allosteric Enzyme by Domain Insertion. **Journal of Molecular Biology**, v. 336, n. 1, p. 263–273, fev. 2004.

- HANSON, J. A.; DUDERSTADT, K.; WATKINS, L. P., BHATTACHARYYA, S.; BROKAW, J.; CHU, J. W.; YANG, H. Illuminating the mechanistic roles of enzyme conformational dynamics. **Proceedings of the National Academy of Sciences** vol. 104, p. 18055-18060, 2007.
- HILSER, V. J.; WRABL, J. O.; MOTLAGH, H. N. Structural and Energetic Basis of Allostery. **Annual Review of Biophysics**, v. 41, n. 1, p. 585–609, 9 jun. 2012.
- HOUGH, D. W.; DANSON, M. J. Extremozymes. **Current Opinion in Chemical Biology**, v. 3, n. 1, p. 39–46, fev. 1999.
- JAFFE, E. K. Morpheesins – a new structural paradigm for allosteric regulation. **Trends in Biochemical Sciences**, v. 30, n. 9, p. 490–497, set. 2005.
- JEFFREY, G. A.; SAENGER, W. **Hydrogen Bonding in Biological Structures**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991.
- KELLEY, L. A.; STERNBERG, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. **Nature Protocols**, v. 4, n. 3, p. 363–371, fev. 2009.
- KOSHLAND, D. E.; NÉMETHY, G.; FILMER, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. **Biochemistry**, v. 5, n. 1, p. 365–385, jan. 1966.
- KUMAR, S.; NUSSINOV, R. How do thermophilic proteins deal with heat? **Cellular and Molecular Life Sciences**, v. 58, n. 9, p. 1216–1233, ago. 2001.
- LEVINE, M.; WEINSTEIN, H. AIM for Allostery: Using the Ising Model to Understand Information Processing and Transmission in Allosteric Biomolecular Systems. **Entropy**, v. 17, n. 5, p. 2895–2918, 7 maio 2015.
- MONOD, J.; CHANGEUX, J.-P.; JACOB, F. Allosteric proteins and cellular control systems. **Journal of Molecular Biology**, v. 6, n. 4, p. 306–329, abr. 1963.
- MONOD, J.; WYMAN, J.; CHANGEUX, J. P. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. **Journal of molecular biology**, v. 12, p. 88–118, maio 1965.
- MOTLAGH, H. N. et al. The ensemble nature of allostery. **Nature**, v. 508, n. 7496, p. 331–339, 16 abr. 2014.
- MURPHY, K. P. **Protein Structure, Stability, and Folding**. New Jersey: Humana Press, 2001. v. 168
- NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry, Fourth Edition**. Fourth Edi ed. [s.l: s.n.].

PERLAND, E.; FREDRIKSSON, R. Classification Systems of Secondary Active Transporters. **Trends in pharmacological sciences**, v. 38, n. 3, p. 305–315, mar. 2017.

PRESLAND, R. B.; DALE, B. A. Epithelial Structural Proteins of the Skin and Oral Cavity: Function in Health and Disease. **Critical Reviews in Oral Biology & Medicine**, v. 11, n. 4, p. 383–408, out. 2000.

SCHWAB, A. et al. Mechanism of Allosteric Inhibition of the Enzyme IspD by Three Different Classes of Ligands. **ACS chemical biology**, v. 12, n. 8, p. 2132–2138, 18 ago. 2017.

SHOICHET, B. K. et al. A relationship between protein stability and protein function. **Proceedings of the National Academy of Sciences**, v. 92, n. 2, p. 452–456, 17 jan. 1995.

SOARES, T. A.; LINS, R. D. Ribozimas: Nem Toda Enzima é uma Proteína. **Química Nova**, v. 18, n. 4, p. 375–378, 1995.

SRINIVASAN, B. et al. Allosteric regulation and substrate activation in cytosolic nucleotidase II from Legionella pneumophila. **FEBS Journal**, v. 281, n. 6, p. 1613–1628, mar. 2014.

TANFORD, C. Protein denaturation. **Advances in protein chemistry**, v. 23, p. 121–282, 1968.

VILLAVERDE, A. Allosteric enzymes as biosensors for molecular diagnosis. **FEBS Letters**, v. 554, n. 1–2, p. 169–172, 6 nov. 2003.

VOET, D.; VOET, J. G.; PRATT, C. W. **Fundamentals of biochemistry**. New York: Wiley, 1999.

WALLQVIST, A.; COVELL, D. G.; THIRUMALAI, D. Hydrophobic Interactions in Aqueous Urea Solutions with Implications for the Mechanism of Protein Denaturation. **Journal of the American Chemical Society**, v. 120, n. 2, p. 427–428, jan. 1998.

WANG, L.-B. et al. A New Nanobiocatalytic System Based on Allosteric Effect with Dramatically Enhanced Enzymatic Performance. **Journal of the American Chemical Society**, v. 135, n. 4, p. 1272–1275, 30 jan. 2013.

WHISSTOCK, J. C.; LESK, A. M. Prediction of protein function from protein sequence and structure. **Quarterly Reviews of Biophysics**, v. 36, n. 3, p. 307–340, ago. 2003.

WYMAN, J.; ALLEN, D.W. The problem of the heme interactions in hemoglobin and the basis of the Bohr effect. **Journal of Polymer Science**, v. 7, p. 499-518, 1951.

WYMAN, J. Heme proteins. **Advances in Protein Chemistry**, vol. 4, p. 407-531, 1948.

YANG, A.-S.; HONIG, B. On the pH Dependence of Protein Stability. **Journal of Molecular Biology**, v. 231, n. 2, p. 459–474, maio 1993.

ZIEGLER, D. M. Role of Reversible Oxidation-Reduction of Enzyme Thiols-Disulfides in Metabolic Regulation. **Annual Review of Biochemistry**, v. 54, n. 1, p. 305–329, jun. 1985.

---

## Capítulo 3

### Xilanases – Histórico e Aplicações

As enzimas xilanases foram descritas pela primeira vez em 1955 como pentosanases ou xilo-dextrinases e seis anos depois foram reconhecidas pela União Internacional de Bioquímica e Biologia Molecular (IUBMB) com o código EC 3.2.1.8 (HOWARD; JONES; PURDOM, 1960; WHISTLER; MASAK, 1955). São produzidas majoritariamente por fungos e bactérias mas também já foram obtidas a partir de protozoários, algas, gastrópodes e artrópodes (DEKKER; RICHARDS, 1976; PRADE, 1996). Essa classe de enzima catalisa a formação de unidades de xilose a partir da quebra do substrato polimérico presente na parede celular das plantas (PAËS; BERRIN; BEAUGRAND, 2012). Essa quebra produz xilooligômeros com cadeias de diferentes tamanhos. Elas existem sob diferentes isoformas, possuem múltiplos mecanismos de ação e variadas propriedades físico-químicas.

Apesar do seu nome oficial ser endo-1,4- $\beta$ -xilanase, essa classe de enzimas possui diversos sinônimos sendo reconhecida também como endoxilanase, 1,4- $\beta$ -DD-xilana-xilana-hidrolase, endo-1,4- $\beta$ -DD-xilanase,  $\beta$ -1,4-xilanase,  $\beta$ -xilanase ou simplesmente xilanase (COLLINS; GERDAY; FELLER, 2005a). Em geral, os nomes trazem a identificação da reação e/ou a posição da ligação que é quebrada/formada por determinada enzima. As xilanases catalisam a hidrólise da xilana, rompendo a ligação entre anéis furanosídicos nas posições 1 e 4 da cadeia principal da xilana.

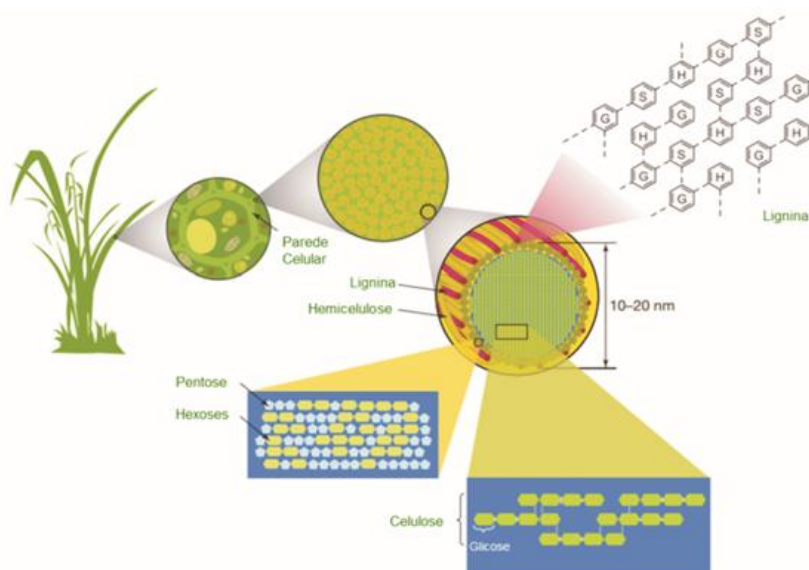
Mesmo com o elevado número de publicações sobre o tema, as xilanases ainda se apresentam como um bom campo de pesquisa com boas oportunidades para o desenvolvimento de processos biotecnológicos devido sua grande flexibilidade e variedade conformacional/estrutural, as quais permitem a criação de

enzimas cada vez mais eficientes a partir de famílias e origens distintas (KULKARNI; SHENDYE; RAO, 1999).

Nas seções a seguir, alguns pontos importantes para o detalhamento desse grupo de enzimas serão discutidos. A seção 3.1 trará informações sobre o substrato no qual atuam as xilanases, posteriormente a classificação e características estruturais dessas enzimas (seção 3.2), informações sobre mutações e xilanases multidomínios (seção 3.3) e ao final do capítulo a apresentação de suas aplicações biotecnológicas, evidenciando o processo de produção do bioetanol desde a biomassa bruta (seção 3.4).

### 3.1 Substrato

O substrato no qual atuam as xilanases é um polissacarídeo extremamente disponível na superfície terrestre, a xilana. Estimativas apontam que a cada três carbonos orgânicos presentes no nosso planeta, um deles está na forma de xilana (PRADE, 1996). Essa grande disponibilidade vem do fato desse substrato ser o maior constituinte da hemicelulose, que juntamente à lignina e celulose compõem a parede celular das plantas (geralmente numa proporção de 3:3:4, respectivamente) como mostrado na Figura 3.1 (CHEN, 2014).



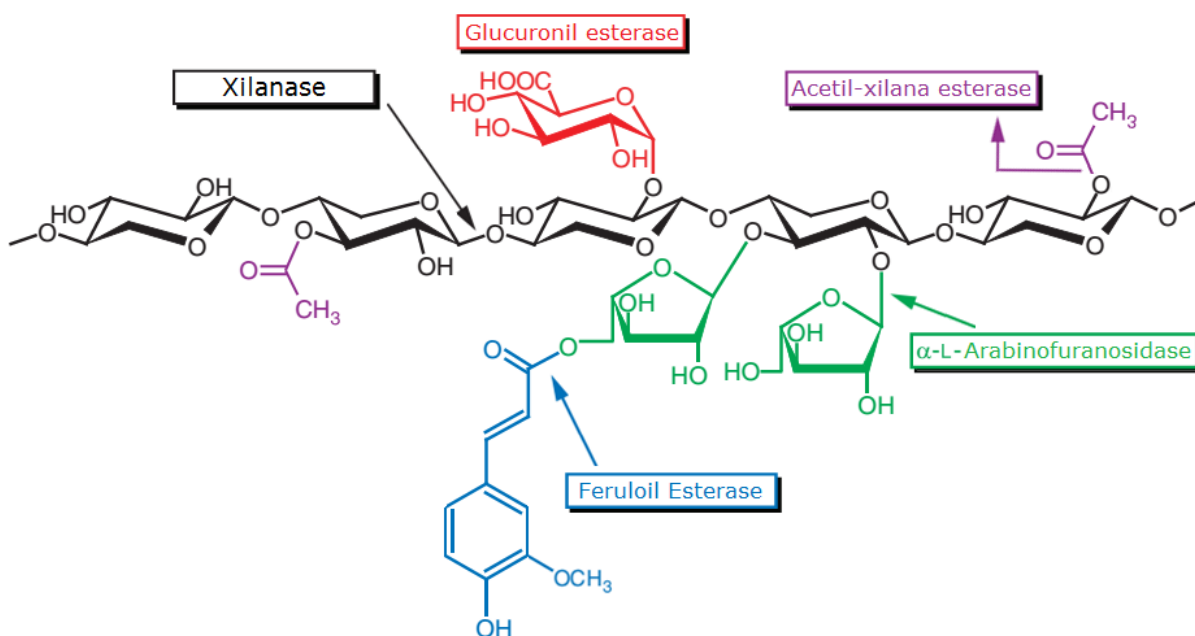
**Figura 3.1** – Estrutura interna da parede celular das plantas evidenciando seus componentes: lignina, hemicelulose e celulose. Modificado de (RUBIN, 2008).

A hemicelulose pode ser conceituada como uma grande rede de polissacarídeos distintos que ligam-se covalentemente entre si e às fibras de celulose (ASPINALL, 1959; BEG et al., 2001; EBRINGEROVÁ; HEINZE, 2000). Como pode-se ver na Figura 3.1, sua estrutura é bem complexa e não possui um padrão específico, podendo ter vários anéis de cinco membros, como unidades de xilose e arabinose, ou de seis membros, como as unidades de glicose e galactose (GÍRIO et al., 2010; SHALLOM; SHOHAM, 2003). Dentre esses vários tipos de heteropolímeros está a xilana. Outros exemplos são as xiloglicanas (polímero heterogêneo de D-xilose e D-glicose), arabinoxilanas (polímero heterogêneo de arabinose e D-xilose) e glicomananas (polímero heterogêneo de D-manose e D-glicose) (LIAB et al., 2000; SAHA, 2003). A lignina por sua vez é amorfa e não possui estrutura bem definida sendo ainda mais complexa que a hemicelulose (CHEN, 2014). Ao contrário delas, a celulose é a parte mais interna das paredes celulares das plantas e tem um empacotamento das cadeias de glicose bem eficiente, garantindo assim, a rigidez necessária à espécie (CHEN, 2014). Mais dados sobre a lignina e celulose podem ser encontrados nas referências a seguir (BAYER et al., 1998; BOERJAN; RALPH; BAUCHER, 2003; CHEN, 2014; KLEMM et al., 2005; LI; PU; RAGAUSKAS, 2016).

A xilana, presente na hemicelulose, é formada por uma cadeia principal de unidades de D-xilopiranosose através de ligações  $\beta$ -1,4 (BEG et al., 2001; SCHELLER; ULVSKOV, 2010). Monossacarídeos e pequenos oligossacarídeos se ligam aleatoriamente a essa cadeia maior em diferentes proporções para formar a estrutura heterogênea da xilana. Esses grupos laterais vêm sendo identificados desde o século passado e os principais são os glicuronopiranosil, 4-O-metil-D-glucuronopiranosil,  $\alpha$ -L-arabinofuranosil, acetil, feruloil ou *p*-cumaroil (DODD; CANN, 2009; SUBRAMANIYAN; PREMA, 2002). A ocorrência de cada um depende da espécie que se deseja analisar (COLLINS; GERDAY; FELLER, 2005a).

Devido sua heterogeneidade e a especificidade enzimática (discutida no capítulo anterior), a hemicelulose precisa de várias enzimas diferentes para ser totalmente quebrada em unidades menores que poderão servir como alimento de micro-organismos (GONÇALVES et al., 2015). Enquanto isso, a celulose, por ser homogênea estruturalmente, necessita basicamente de apenas de uma enzima, a celulase, para realizar a clivagem dos monômeros de D-glicose. As enzimas que

participam do processo hemicelulolítico são conhecidas como enzimas acessórias, já que auxiliam no processo principal de degradação da parede celular das plantas, no qual é feito pelas celulasas sobre a celulose. A principal delas é a xilanase, que quebra a cadeia principal desse polissacarídeo em diferentes posições da cadeia. As demais são enzimas acessórias que levam os nomes dos grupos que clivam, ou seja,  $\alpha$ -L-arabinofuranosídeses (EC 3.2.1.55), glucuronil esterases (EC 3.2.1.139), acetilxilanoesterases (EC 3.1.1.72), feruloil esterases (EC 3.1.1.73) e *p*-cumaroil esterases (EC 3.1.1.-), Fig. 3.2 (COLLINS; GERDAY; FELLER, 2005a; MOREIRA; FILHO, 2016).



**Figura 3.2** – Estrutura representativa de uma parte da xilana. Cadeia principal formada por monômeros de D-xilopiranosose (preto) e possíveis cadeias laterais de acetato (roxo), arabinofuranose (verde), ácido ferúlico (azul) e D-glucuronopiranosil (vermelho) com suas respectivas enzimas destacadas. Modificado de (DODD; CANN, 2009).

### 3.2 Classificação e características estruturais

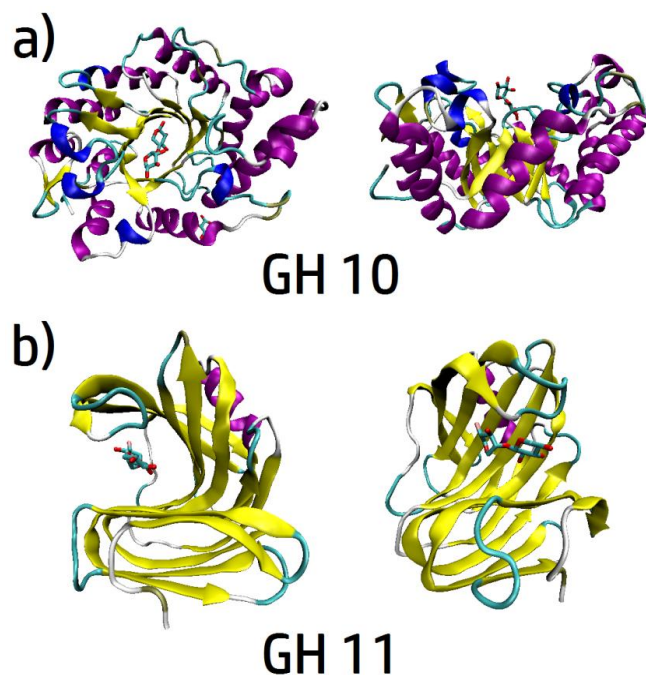
Alguns sistemas de classificação de xilanases já foram propostos levando em consideração, por exemplo, o peso molecular e o ponto isoelétrico de cada enzima de forma independente (WONG; TAN; SADDLER, 1988). Essas classificações não compreendiam bem as divisões existentes entre as xilanases que evolutivamente se desenvolveram de forma muito particular. A saída foi criar um conceito mais amplo



utilizando determinados parâmetros pelos quais se agruparia todas as glicosidades. Essa classe de enzimas, também conhecida como glicosil hidrolases (GHs), catalisa a quebra de ligações glicosídicas, ou seja, é um grupo mais amplo que compreende todas as xilanases (HENRISSAT et al., 1989). O código de cada uma revela isso, sendo EC 3.2.1.x e EC 3.2.1.8 para GHs e xilanases, respectivamente. Esse formato de classificação é baseado na comparação entre a estrutura primária do sítio catalítico e nos grupos de enzimas com sequência similar (HENRISSAT; COUTINHO, 2001). Essa classificação existe até hoje com registros de mais de 100 famílias. Entretanto, devido a evolução e grande variedade conhecida atualmente, outra forma de agrupar as xilanases foi proposta baseando-se na similaridade de sua estrutura tridimensional (HENRISSAT; BAIROCH, 1993). Nesse modo, as centenas de glicosidades foram subdivididas em grandes grupos, os quais também são chamados de clãs (ou famílias). Somente alguns desses possuem atividade xilanolítica, basicamente as famílias 10 e 11, também conhecidas como GH 10 e GH 11 (BOURNE; HENRISSAT, 2001; CLAEYSSSENS; HENRISSAT, 1992). Outras famílias apresentam enzimas com atividades xilanolíticas mas são casos isolados que não chegam a uma dezena de exemplos (como as famílias GH 5, GH 7, GH 8 e GH 43), enquanto as famílias GH 10 e GH 11 já somam mais de uma centena de casos cada (COLLINS; GERDAY; FELLER, 2005a).

A família GH 10 tem a característica de ter enzimas com alta massa molecular, baixo ponto isoelétrico e estrutura tridimensional em formato de metade de uma esfera, ou seja, com uma superfície externa maior que a interna na qual se liga o substrato (Fig. 3.3a). Já a família GH 11 apresenta enzimas com peso molecular baixo e alto ponto isoelétrico que estruturalmente se assemelham à mão direita (Fig. 3.3b) (HAKULINEN et al., 2003; LO LEGGIO et al., 1999; PELL et al., 2004).

A família GH 11 em especial possui maior especificidade em relação ao substrato xilanolítico quando comparada às demais. Enquanto que outras famílias, como a GH 10, possui enzimas com atividades para outros substratos semelhantes, as enzimas pertencentes à família GH 11 são específicas para xilana (BIELY et al., 1997).



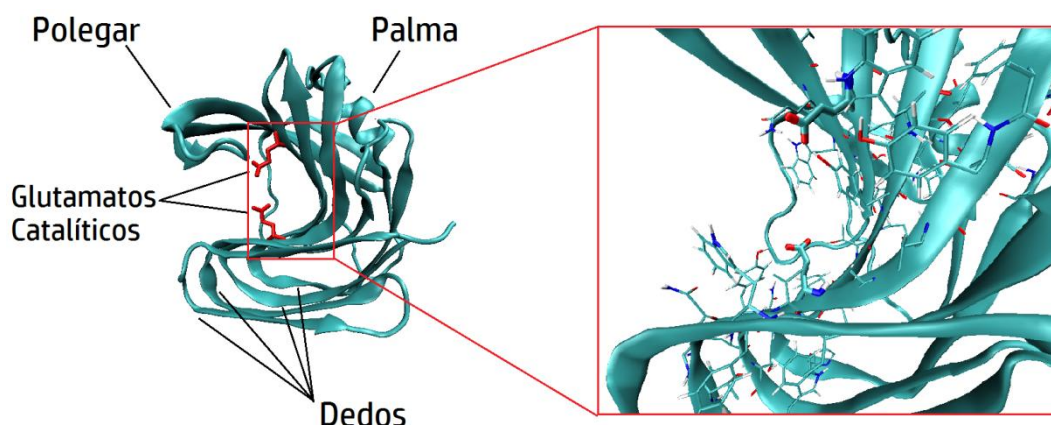
**Figura 3.3** – Estrutura representativa de xilanases a) GH 10 e b) GH 11. O substrato está evidenciado próximo às cavidades catalíticas.  $\alpha$ -hélices (roxo) e  $\beta$ -folhas (amarelo) estão apresentadas em destaque. Imagens produzidas através do *software* Visual Molecular Dynamics 1.9.2.

Sua estrutura no formato de mão direita evidencia a região dos quatro dedos, da palma e do polegar, como exemplificado na Figura 3.4 (TÖRRÖNEN; HARKKI; ROUVINEN, 1994). A cavidade formada entre os dedos e o polegar varia de acordo com o movimento abre-fecha da enzima (DE LEMOS ESTEVES et al., 2004; MUILU et al., 1998). Esse movimento é característico das xilanases dessa família. O substrato se encaixa perfeitamente e passa pelo mecanismo de substituição nucleofílica de primeira ordem com duplo deslocamento através de dois resíduos de glutamato (carregados negativamente em pH neutro) presentes na parte interna da palma que funcionam como nucleófilos (LAWSON; WAKARCHUK; WITHERS, 1996, 1997; MOREIRA; FILHO, 2016).

### 3.3 Mutações e xilanases multidomínio

As xilanases, principalmente da família GH 11, possuem uma grande versatilidade para processos biotecnológicos. Muitas pesquisas vêm sendo conduzidas para que essas enzimas sejam cada vez mais eficientes e resistentes termicamente. Essas modificações geralmente têm fins específicos e além da

estabilidade térmica e maior atividade xilanolítica, podem também aumentar a resistência para pHs extremos, diminuir a competição entre substratos semelhantes no sítio catalítico ou aumentar a sinergia entre enzimas que trabalham simultaneamente no mesmo meio (COLLINS; GERDAY; FELLER, 2005b; PAËS; BERRIN; BEAUGRAND, 2012; PRADE, 1996). Para tal, a utilização da engenharia de proteínas tem papel de destaque no que se refere à inserção de domínios (módulos) e mutações.



**Figura 3.4** – Estrutura representativa de uma xilanase GH 11 (MURAKAMI et al., 2005). Evidenciando regiões importantes e os resíduos catalíticos (vermelho). Imagens produzidas através do *software* Visual Molecular Dynamics 1.9.2.

Os domínios são partes funcionais de proteínas que, evolutiva ou sinteticamente, foram criadas para realizar uma determinada função. Muitas xilanases conhecidas já possuem domínios associados para melhorar determinada propriedade. A engenharia de enzimas multidomínios (enzimas capazes de catalisar duas ou mais reações) é uma promissora estratégia para reduzir os custos enzimáticos em processos pois propriedades catalíticas múltiplas estão inseridas em uma única enzima. Essas enzimas multidomínios ou enzimas híbridas, formadas pela fusão de dois ou mais domínios, são denominadas quimeras. Geralmente, as enzimas quiméricas apresentam propriedades catalíticas melhoradas quando comparada com as de suas enzimas individuais (FONSECA-MALDONADO et al., 2014; RIBEIRO et al., 2016). Quimeras são construídas por evolução dirigida de tal modo que, duas proteínas existentes, uma com a função de reconhecimento do sinal (proteína A) e outra com uma função a ser modulada (proteína B) (Cap. 6), são fundidas de forma que o reconhecimento do sinal pelo domínio A é transmitido para

o domínio B, por mudanças conformacionais no domínio A, e assim, a atividade do domínio B é afetada, ou seja, apresentam efeito alostérico (apresentado no capítulo anterior).

Mutações também são muito comuns nessa classe de enzimas. Uma publicação originada desta tese utilizou a substituição de aminoácidos, ou seja, a mudança de polaridade ou tamanho das cadeias laterais, como agente transformador de determinado parâmetro (SILVA et al., 2017) (Cap. 5). A maioria das mutações se concentra no aumento da termoestabilidade e da atividade através da modificação de resíduos em regiões dos dedos e da cavidade catalítica, respectivamente.

### 3.4 Aplicações

Devido suas propriedades físicas e especificidade, as aplicações biotecnológicas dessas enzimas varrem diversas áreas. São usadas no processo de branqueamento do papel, na manipulação de massas comestíveis, na purificação e tratamento de sucos, na suplementação de animais, e, principalmente, na produção de etanol lignocelulósico (BEG et al., 2001; BHAT, 2000; COLLINS; GERDAY; FELLER, 2005a; OAKLEY et al., 2003; SUBRAMANIYAN; PREMA, 2002).

#### *3.4.1 Produção do bioetanol*

Nos dias atuais, o principal interesse está em utilizar essas enzimas na degradação da celulose em açúcares fermentáveis para subsequente produção de etanol, por exemplo, usando como biomassa o bagaço da cana-de-açúcar ou a celulose da madeira (SUN; CHENG, 2002). Como já comentado anteriormente as fibras de celulose encontram-se estabilizadas por interações com a lignina e a hemicelulose (CHEN, 2014). Logo, há um entrave tecnológico para que essas cadeias sejam separadas e então processadas de maneira adequada. Então, a ideia é que diversas enzimas atuem sinergicamente para quebra dessas cadeias dispensando o uso de agentes inorgânicos agressivos ambientalmente (RYU; KARIM, 2011). Essa produção, desde a biomassa até o etanol de segunda geração

pode ser resumido em três etapas, o pré-tratamento, a hidrólise enzimática e a fermentação (POTHIRAJ; KANMANI; BALAJI, 2006).

No pré-tratamento a biomassa deve ser processada, por exemplo, se a biomassa é a madeira, ela deve ser cortada em pedaços menores para facilitar o seu transporte e também para aumentar a área superficial para as reações químicas. Também deve receber um tratamento para enfraquecer ou romper suas fibras, por exemplo, através de líquidos iônicos (ELGHARBAWY et al., 2016).

Posteriormente, as fibras pré-tratadas são submetidas a uma solução enzimática para quebrar a celulose em açúcares simples, como a glicose. Esse processo é chamado de hidrólise enzimática. Essa solução, ou “coquetel” enzimático como é comumente chamada, é composta por diferentes enzimas celulasas e hemicelulasas. As hemicelulasas são as primeiras enzimas a serem adicionadas. Elas facilitam o acesso da enzima celulase até a celulose. Um dos tipos são as endocelulasas ou  $\beta$ -1,4-endoglucanases (EC 3.2.1.4) que quebram as ligações internas para romper a estrutura cristalina da celulose e expor as cadeias polissacarídicas individuais de celulose (DODD; CANN, 2009; RYU; KARIM, 2011). Já as exocelulasas (EC 3.2.1.91) clivam duas ou quatro unidades do final das cadeias expostas pelas endocelulasas, resultando em tetrassacarídeos ou dissacarídeos, como a celobiose e as  $\beta$ -glucosidases (EC 3.2.1.21) hidrolisam os produtos das exocelulasas (celobioses) em monossacarídeos individuais (glicose).

Ao final dá-se o processo de fermentação, ou seja, a solução de glicose é usada para alimentar microorganismos, tais como leveduras, que consumirão a glicose produzindo etanol como um subproduto do seu metabolismo (OLSSON; HAHN-HÄGERDAL, 1996). Essa etapa não é diferente do processo fermentativo que transforma uvas em vinho. Posteriormente, o etanol deve ser destilado para que a sua concentração aumente, podendo atingir 97%, de modo que outras técnicas devem ser aplicadas para remover o teor aquoso restante (BRETHAUER; WYMAN, 2010). O etanol produzido por este processo recebe o nome de bioetanol, etanol celulósico ou etanol de 2º geração.

## Referências - Capítulo 3

- ASPINALL, G. O. Structural chemistry of the hemicelluloses. **Advances in carbohydrate chemistry**, v. 14, p. 429–68, 1959.
- BAYER, E. A. et al. Cellulose, cellulases and cellulosomes. **Current opinion in structural biology**, v. 8, n. 5, p. 548–57, out. 1998.
- BEG, Q. K. et al. Microbial xylanases and their industrial applications: a review. **Applied microbiology and biotechnology**, v. 56, n. 3–4, p. 326–38, ago. 2001a.
- BHAT, M. K. Cellulases and related enzymes in biotechnology. **Biotechnology advances**, v. 18, n. 5, p. 355–83, ago. 2000.
- BIELY, P. et al. Endo- $\beta$ -1,4-xylanase families: differences in catalytic properties. **Journal of Biotechnology**, v. 57, n. 1–3, p. 151–166, set. 1997.
- BOERJAN, W.; RALPH, J.; BAUCHER, M. Lignin biosynthesis. **Annual review of plant biology**, v. 54, p. 519–46, 2003.
- BOURNE, Y.; HENRISSAT, B. Glycoside hydrolases and glycosyltransferases: families and functional modules. **Current opinion in structural biology**, v. 11, n. 5, p. 593–600, out. 2001.
- BRETHAUER, S.; WYMAN, C. E. Review: Continuous hydrolysis and fermentation for cellulosic ethanol production. **Bioresource Technology**, v. 101, n. 13, p. 4862–4874, jul. 2010.
- CHEN, H. Chemical Composition and Structure of Natural Lignocellulose. In: **Biotechnology of Lignocellulose: Theory and Practice**. Dordrecht: Springer Netherlands, 2014. p. 25–71.
- CLAEYSSSENS, M.; HENRISSAT, B. Specificity mapping of cellulolytic enzymes: classification into families of structurally related proteins confirmed by biochemical analysis. **Protein science: a publication of the Protein Society**, v. 1, n. 10, p. 1293–7, out. 1992.
- COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanase families and extremophilic xylanases. **FEMS microbiology reviews**, v. 29, n. 1, p. 3–23, jan. 2005a.
- COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanase families and extremophilic xylanases. **FEMS Microbiology Reviews**, v. 29, n. 1, p. 3–23, jan. 2005b.
- DE LEMOS ESTEVES, F. et al. Acidophilic adaptation of family 11 endo- $\beta$ -1,4-

xylanases: Modeling and mutational analysis. **Protein Science**, v. 13, n. 5, p. 1209–1218, maio 2004.

DEKKER, R. F. H.; RICHARDS, G. N. Hemicellulases: Their Occurrence, Purification, Properties, and Mode of Action. **Advances in Carbohydrate Chemistry and Biochemistry**, v. 32, p. 277–352, 1976.

DODD, D.; CANN, I. K. O. Enzymatic deconstruction of xylan for biofuel production. **GCB Bioenergy**, v. 1, n. 1, p. 2–17, fev. 2009.

EBRINGEROVÁ, A.; HEINZE, T. Xylan and xylan derivatives - biopolymers with valuable properties, 1. Naturally occurring xyans structures, isolation procedures and properties. **Macromolecular Rapid Communications**, v. 21, n. 9, p. 542–556, 1 jun. 2000.

ELGHARBAWY, A. A. et al. Ionic liquid pretreatment as emerging approaches for enhanced enzymatic hydrolysis of lignocellulosic biomass. **Biochemical Engineering Journal**, v. 109, p. 252–267, maio 2016.

FONSECA-MALDONADO, R. et al. Synergistic action of co-expressed xylanase/laccase mixtures against milled sugar cane bagasse. **Process Biochemistry**, v. 49, n. 7, p. 1152–1161, jul. 2014.

GÍRIO, F. M. et al. Hemicelluloses for fuel ethanol: A review. **Bioresource Technology**, v. 101, n. 13, p. 4775–4800, jul. 2010.

GONÇALVES, G. A. L. et al. Synergistic effect and application of xylanases as accessory enzymes to enhance the hydrolysis of pretreated bagasse. **Enzyme and microbial technology**, v. 72, p. 16–24, maio 2015.

HAKULINEN, N. et al. Three-dimensional structures of thermophilic beta-1,4-xylanases from *Chaetomium thermophilum* and *Nonomuraea flexuosa*. Comparison of twelve xylanases in relation to their thermal stability. **European journal of biochemistry**, v. 270, n. 7, p. 1399–412, abr. 2003.

HENRISSAT, B. et al. Cellulase families revealed by hydrophobic cluster analysis. **Gene**, v. 81, n. 1, p. 83–95, 1 set. 1989.

HENRISSAT, B.; BAIROCH, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. **The Biochemical journal**, v. 293 ( Pt 3, p. 781–8, 1 ago. 1993.

HENRISSAT, B.; COUTINHO, P. M. Classification of glycoside hydrolases and glycosyltransferases from hyperthermophiles. **Methods in enzymology**, v. 330, p.

183–201, 2001.

HOWARD, B. H.; JONES, G.; PURDOM, M. R. The pentosanases of some rumen bacteria. **The Biochemical journal**, v. 74, p. 173–80, jan. 1960.

KLEMM, D. et al. Cellulose: fascinating biopolymer and sustainable raw material. **Angewandte Chemie (International ed. in English)**, v. 44, n. 22, p. 3358–93, 30 maio 2005.

KULKARNI, N.; SHENDYE, A.; RAO, M. Molecular and biotechnological aspects of xylanases. **FEMS microbiology reviews**, v. 23, n. 4, p. 411–56, jul. 1999.

LAWSON, S. L.; WAKARCHUK, W. W.; WITHERS, S. G. Effects of both Shortening and Lengthening the Active Site Nucleophile of *Bacillus circulans* Xylanase on Catalytic Activity †. **Biochemistry**, v. 35, n. 31, p. 10110–10118, jan. 1996.

LAWSON, S. L.; WAKARCHUK, W. W.; WITHERS, S. G. Positioning the acid/base catalyst in a glycosidase: studies with *Bacillus circulans* xylanase. **Biochemistry**, v. 36, n. 8, p. 2257–65, 25 fev. 1997.

LI, M.; PU, Y.; RAGAUSKAS, A. J. Current Understanding of the Correlation of Lignin Structure with Biomass Recalcitrance. **Frontiers in Chemistry**, v. 4, 18 nov. 2016.

LIAB et al. Relationships between activities of xylanases and xylan structures. **Enzyme and microbial technology**, v. 27, n. 1–2, p. 89–94, 1 jul. 2000.

LO LEGGIO, L. et al. High resolution structure and sequence of *T. aurantiacus* xylanase I: implications for the evolution of thermostability in family 10 xylanases and enzymes with (beta)alpha-barrel architecture. **Proteins**, v. 36, n. 3, p. 295–306, 15 ago. 1999.

MOREIRA, L. R. S.; FILHO, E. X. F. Insights into the mechanism of enzymatic hydrolysis of xylan. **Applied Microbiology and Biotechnology**, v. 100, n. 12, p. 5205–5214, 25 jun. 2016.

MUILU, J. et al. Functional conformational changes of endo-1,4-xylanase II from *Trichoderma reesei*: a molecular dynamics study. **Proteins**, v. 31, n. 4, p. 434–44, 1 jun. 1998.

MURAKAMI, M. T. et al. Correlation of temperature induced conformation change with optimum catalytic activity in the recombinant G/11 xylanase A from *Bacillus subtilis* strain 168 (1A1). **FEBS Letters**, v. 579, n. 28, p. 6505–6510, 21 nov. 2005.

OAKLEY, A. J. et al. Characterization of a family 11 xylanase from *Bacillus subtilis* B230 used for paper bleaching. **Acta crystallographica. Section D, Biological**



**crystallography**, v. 59, n. Pt 4, p. 627–36, abr. 2003.

OLSSON, L.; HAHN-HÄGERDAL, B. Fermentation of lignocellulosic hydrolysates for ethanol production. **Enzyme and Microbial Technology**, v. 18, n. 5, p. 312–331, abr. 1996.

PAËS, G.; BERRIN, J.-G.; BEAUGRAND, J. GH11 xylanases: Structure/function/properties relationships and applications. **Biotechnology advances**, v. 30, n. 3, p. 564–92, 2012.

PELL, G. et al. Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. **The Journal of biological chemistry**, v. 279, n. 12, p. 11777–88, 19 mar. 2004.

POTHIRAJ, C.; KANMANI, P.; BALAJI, P. Bioconversion of Lignocellulose Materials. **Mycobiology**, v. 34, n. 4, p. 159, 2006.

PRADE, R. A. Xylanases: from biology to biotechnology. **Biotechnology & genetic engineering reviews**, v. 13, p. 101–31, jan. 1996.

RIBEIRO, L. F. et al. A xylose-stimulated xylanase-xylose binding protein chimera created by random nonhomologous recombination. **Biotechnology for biofuels**, v. 9, p. 119, 2016.

RYU, S.; KARIM, M. N. A whole cell biocatalyst for cellulosic ethanol production from dilute acid-pretreated corn stover hydrolyzates. **Applied microbiology and biotechnology**, v. 91, n. 3, p. 529–42, ago. 2011.

SAHA, B. C. Hemicellulose bioconversion. **Journal of industrial microbiology & biotechnology**, v. 30, n. 5, p. 279–91, maio 2003.

SCHELLER, H. V.; ULVSKOV, P. Hemicelluloses. **Annual Review of Plant Biology**, v. 61, n. 1, p. 263–289, 2 jun. 2010.

SHALLOM, D.; SHOHAM, Y. Microbial hemicellulases. **Current opinion in microbiology**, v. 6, n. 3, p. 219–28, jun. 2003.

SILVA, S. B. et al. The role of local residue environmental changes in thermostable mutants of the GH11 xylanase from *Bacillus subtilis*. **International Journal of Biological Macromolecules**, v. 97, p. 574–584, abr. 2017.

SUBRAMANIYAN, S.; PREMA, P. Biotechnology of microbial xylanases: enzymology, molecular biology, and application. **Critical reviews in biotechnology**, v. 22, n. 1, p. 33–64, jan. 2002.

SUN, Y.; CHENG, J. Hydrolysis of lignocellulosic materials for ethanol production: a review. **Bioresource Technology**, v. 83, n. 1, p. 1–11, maio 2002.

TÖRRÖNEN, A.; HARKKI, A.; ROUVINEN, J. Three-dimensional structure of endo-1,4-beta-xylanase II from *Trichoderma reesei*: two conformational states in the active site. **The EMBO Journal**, v. 13, n. 11, p. 2493–2501, 1 jun. 1994.

WHISTLER, R. L.; MASAK, E. Enzymatic Hydrolysis of Xylan 1. **Journal of the American Chemical Society**, v. 77, n. 5, p. 1241–1243, mar. 1955.

WONG, K. K.; TAN, L. U.; SADDLER, J. N. Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. **Microbiological reviews**, v. 52, n. 3, p. 305–17, set. 1988.

---

## Capítulo 4

### Fundamentos da Dinâmica Molecular

As simulações computacionais de sistemas biológicos utilizando a dinâmica molecular vem sendo uma das primeiras opções nas últimas décadas quando necessita-se obter dados teóricos de movimentação estrutural e interações intra e intermoleculares de biomoléculas. (KARPLUS; KURIYAN, 2005; KARPLUS; MCCAMMON, 2002). A dinâmica molecular é uma técnica computacional que simula a evolução temporal de átomos e moléculas através da ação de forças provenientes de um potencial de interação responsável pela modelagem do sistema químico específico (FRENKEL; SMIT, 2001).

Devido à limitações de *hardware* e *software*, os tempos simulados de sistemas biológicos, como as proteínas ou porções do DNA, são na escala de nano ou microssegundos (BERNARDI; MELO; SCHULTEN, 2015). Apesar de parecer um tempo muito curto, os resultados dessas simulações podem fornecer dados confiáveis já que os eventos de interesse possuem a mesma escala temporal. Assim como as demais técnicas computacionais, a qualidade dos resultados de uma dinâmica molecular clássica, doravante DM, tem uma relação direta com o custo computacional envolvido e os parâmetros de cálculo escolhidos para descrição do sistema. Em geral, essa balança precisa ser bem ajustada pelo usuário para que se consiga obter os melhores resultados com os equipamentos disponíveis (KARPLUS; MCCAMMON, 2002; SHAW et al., 2010).

O bom uso dessa técnica pode trazer benefícios como o suporte ou até mesmo a substituição de recursos experimentais os quais carregam consigo riscos ambientais e ao usuário somados aos custos de operação elevados. Além disso, pode eliminar metodologias ineficientes como as de tentativa-erro *in vitro* e/ou *in vivo*

através de testes *in silico* dos mesmos sistemas. A química computacional também configura-se como uma ferramenta importante no âmbito da química limpa pois pode diminuir ou eliminar o consumo de reagentes, principalmente solventes nocivos, por não proporcionar a geração de resíduos químicos e por evitar possíveis acidentes como vazamentos e/ou incêndios. Em contrapartida, possui uma alta demanda energética para manter as máquinas funcionando por longos períodos que por vezes chega a meses de trabalho ininterrupto. (ANASTAS; KIRCHHOFF, 2002).

A literatura atual dispõe de várias metodologias para descrição completa e adequada de sistemas moleculares, seja de gases, líquidos ou sólidos. Uma seção especial desses modelos é voltada para sistemas envolvendo proteínas, DNA, carboidratos, lipídios e outras estruturas orgânicas e biológicas (GELPI et al., 2015). Em geral, esses biosistemas são grandes do ponto de vista atômico e demandam grande custo computacional. Logo, precisam ser tratados de forma mais simples, em termos matemáticos, para que sejam estudados em uma escala de tempo adequada (FRENKEL; SMIT, 2001). Os métodos em química computacional podem ser divididos de várias maneiras, dentre elas, os algoritmos que realizam cálculos de propriedades intrínsecas das moléculas (momento dipolo, representação de orbitais, cargas atômicas, dentre outros), e aqueles pelos quais pode-se obter parâmetros de um conjunto de moléculas através da análise da sua movimentação e interação com as demais.

Os cálculos nos quais é levado em conta somente a contribuição potencial podem prever com exatidão propriedades elétricas e estruturais das moléculas através da otimização estrutural, cálculo de frequência vibracional e/ou análises de energia de orbitais (SZABO; OSTLUND, 1996; YOUNG, 2001). Métodos como Hartree-Fock, Teoria do Funcional da Densidade (DFT), Teoria da Perturbação ou métodos semi-empíricos têm sido mais utilizados para tais finalidades com aplicações diversas desde meados do século passado, embora também existam métodos baseados em mecânica clássica (ARGAMAN; MAKOV, 2000; CHUNG et al., 2015; LEVINE, 2009; SLATER, 1951). Inclusive, dois artigos foram publicados durante o período do doutoramento nos quais foram empregados métodos baseados em química quântica. Em um deles o método semiempírico foi empregado na descrição a nível molecular de um complexo metálico de ferro com moléculas de orto-fenantrolina, e no outro, cálculos usando DFT e mecânica molecular foram

empregados na descrição de interações intermoleculares entre ligantes estereoidais e a enzima CYP17 (DE MORAIS et al., 2016; FIRME; MONTEIRO; SILVA, 2017).

Por outro lado, outros métodos consideram a movimentação de centenas ou milhares de moléculas pelo sistema. Neste caso, a contribuição cinética é levada em consideração juntamente com o potencial que as descreve. Para tanto, existem métodos baseados na mecânica molecular, ditos portanto, clássicos, e aqueles que utilizam os preceitos da mecânica quântica em sua descrição, doravante, quânticos. Nos métodos que consideram potenciais clássicos, como a dinâmica clássica, simulações de Monte Carlo e metadinâmica, grandes sistemas podem ser descritos (proteínas, membranas e ácidos nucleicos) apesar de não considerarem a quebra e formação de ligações durante o processo. Algumas possibilidades surgiram nas últimas décadas para que seja possível o tratamento dinâmico de grandes sistemas simultaneamente com a descrição molecular de reações químicas, principalmente no meio biológico. Apesar de caros computacionalmente, parte do sistema de interesse descrito através da mecânica quântica e o restante (maior parte) continua a ser tratado via mecânica clássica. Além desse método híbrido de simulação (FIELD; BASH; KARPLUS, 1990; WARSHEL; LEVITT, 1976), outras opções já vêm sendo utilizadas para deixar o sistema melhor modelado em um tempo inferior, como o uso de constante dielétrica para simular solventes (ROUX; SIMONSON, 1999) ou até mesmo consideração de conjuntos de átomos como sendo um grande grupo atômico para diminuir o custo computacional (*coarse-grained modeling*) (BOND et al., 2007; HADLEY; MCCABE, 2012).

Na presente tese, as xilanases, enzimas que dentre outros processos, participam do processo de produção do bioetanol foram estudadas por DM. Vários fatores contribuíram para essa escolha. O a) tamanho dos sistemas estudados: as xilanases são proteínas com cerca de 185 aminoácidos em sua estrutura primária e com adição de solvente, a condição de simulação teria mais de 80 mil átomos que inviabiliza o tratamento utilizando cálculos baseados em mecânica quântica; b) observação da dinâmica enzimática: as xilanases e a XBP (*xylose binding protein*, tratada no Cap. 6) possuem movimentos característicos de abre-fecha que precisam ser considerados durante a simulação, além disso o efeito alostérico só é revelado pela dinâmica de partes das enzimas, logo, é crucial que o método descreva bem esse processo mesmo em grandes sistemas; c) estudo de interações entre grupos

de átomos: a DM possui a característica de ser baseada na interação entre cargas pontuais de massas correspondentes ao peso atômico e seus resultados de energia de interação (quando o campo de força utilizado está bem parametrizado) são bastante confiáveis e comparáveis com dados experimentais e/ou obtidos por métodos quânticos; d) irrelevância da observação de reações químicas: o foco do trabalho não é buscar superfícies de energia potencial (SEP) que descrevam estados de transição ou intermediários reacionais, logo, é dispensável o uso de métodos baseado em mecânica quântica que descrevam tais eventos já que seu uso seria inviável para sistemas tão grandes como as proteínas.

Portanto, durante esse capítulo, o cerne serão os conceitos e a matemática envolvida exclusivamente na DM (seção 4.1). Posteriormente serão discutidos os parâmetros que compõem os campos de força (seção 4.2) e por fim, as ferramentas de análise disponíveis no âmbito da DM (seção 4.3).

#### 4.1 Conceitos de simulação molecular

As simulações de dinâmica molecular modelam o movimento e interação entre várias partículas dentro de um determinado sistema. Nesse contexto, os átomos, ou moléculas em alguns casos, são considerados como corpos com massa e carga específica que se movimentam regidos por equações de movimento baseadas nas leis da mecânica clássica. Por esse motivo, a técnica é conhecida como dinâmica molecular clássica (ALLEN; TILDESLEY, 1989).

Assim como em um experimento real de bancada, as simulações seguem conceitos semelhantes. Inicialmente devemos preparar as amostras que iremos estudar, posteriormente escolhemos como nossa propriedade de interesse será medida, com qual instrumento isso será feito e por quanto tempo será observado e por fim, coletamos os dados obtidos. No caso, criamos o sistema com  $N$  partículas, resolvemos as equações de movimento por um tempo adequado para que nossa propriedade de interesse não varie com o tempo, indicando, portanto, o equilíbrio desejado.

O Hamiltoniano ( $H$ ) descreve a energia total de um sistema, ou seja, a soma da contribuição cinética e potencial dos átomos e/ou moléculas envolvidos. Logo, a descrição do sistema passará pela seguinte equação:

$$H(x, y, z, p) = K(p) + V(x, y, z) = E_{\text{Total}} \quad (1)$$

onde o  $K(p)$  é o termo descritivo da energia cinética que é função do momento linear ( $p$ ) de cada centro de massa simulado, enquanto que o  $V(x, y, z)$  representará o potencial de interação entre os corpos sendo dependente das coordenadas, em geral, cartesianas, dessas massas.

Estendendo o termo da energia cinética, teremos a seguinte equação:

$$K(p) = \sum_{i=1}^N \sum_{\alpha} p_{i\alpha}^2 / 2m_i \quad (2)$$

na qual o termo  $K(p)$  é um somatório dos componentes de momento linear nos eixos  $x$ ,  $y$  e  $z$ , representando pelo índice  $\alpha$ , para todos os  $N$  corpos do sistema com massa correspondente ( $m_i$ ). Já o termo  $V(x, y, z)$  será melhor discutido na seção 4.2 pois é subdividido nos parâmetros englobados pelos campos de força.

#### 4.1.1 Equações de movimento

Para o cálculo do movimento das massas a serem simuladas, precisamos do conceito clássico da segunda Lei de Newton escrita na forma de equação diferencial:

$$\vec{F}_i = m_i \cdot \vec{a}_i = m_i \cdot \frac{d^2 \vec{r}_i(t)}{dt^2} = - \frac{\partial V[(r)^N]}{\partial \vec{r}_i} \quad (3)$$

onde  $i = 1, 2, \dots, N$  partículas,  $\vec{F}_i$  é a força atuante no átomo  $i$ ,  $m_i$  e  $\vec{a}_i$  sendo sua massa e aceleração respectivamente,  $\vec{r}_i$  a posição da partícula  $i$  no espaço tridimensional e  $V[(r)^N]$  como a energia potencial total do sistema.

Devido o exposto pela Eq. 3, existe uma limitação de tamanho para os sistemas na DM. É inviável nos computadores atuais, o cálculo de um mol de substâncias por exemplo, visto que teríamos que calcular os três componentes ( $x, y$  e  $z$ ) da força para cada uma das  $10^{23}$  partículas, além de suas velocidades a cada

passo. Por isso que os sistemas são pequenos (considerados microscópicos), necessitando de grupos representativos para descrição do comportamento total, os chamados *ensembles*.

Para que uma DM seja feita, não basta apenas sabermos a força atuante em cada átomo e/ou molécula ou suas coordenadas, é necessário que integremos as equações de movimento acima a fim de obtermos as trajetórias ao longo de determinado espaço de tempo. Mais notadamente destacam-se dois algoritmos que realizam essa tarefa, o *Leap-Frog* e *Verlet* (ALLEN; TILDESLEY, 1989; VAN GUNSTEREN; BERENDSEN, 1988).

O *Leap-Frog* utiliza-se somente da posição atual e da velocidade da partícula num tempo igual a metade de um período de integração para obter uma posição futura dessa mesma partícula. Utilizando as equações abaixo, podemos entender o funcionamento desse algoritmo:

$$\vec{v}\left(t - \frac{\delta t}{2}\right) = \frac{\vec{r}(t) - \vec{r}(t - \delta t)}{\delta t} \quad (4)$$

$$\vec{v}\left(t + \frac{\delta t}{2}\right) = \frac{\vec{r}(t + \delta t) - \vec{r}(t)}{\delta t} \quad (5)$$

As Eqs. 4 e 5 podem ser rearranjadas isolando-se o termo da posição em um tempo futuro  $\vec{r}(t + \delta t)$ ,

$$\vec{r}(t + \delta t) = \left[ \vec{v}\left(t + \frac{\delta t}{2}\right) \cdot \delta t \right] + \vec{r}(t) \quad (6)$$

assim, teremos o conceito que foi descrito mais acima, sendo  $\vec{v}\left(t + \frac{\delta t}{2}\right)$  a velocidade da partícula daqui a meio passo tempo, o  $\delta t$  sendo um passo temporal completo e  $\vec{r}(t)$  como a posição da partícula no tempo atual.

Se existe a informação da localização tridimensional da partícula e pode-se estabelecer o tamanho do passo de integração, só resta obter a velocidade daqui a meio passo temporal e então tem-se a posição da partícula nesse determinado tempo futuro. Para isso podemos usar as expansões de Taylor para um tempo futuro (Eq. 7) e um tempo passado (Eq. 8) a partir dos conhecimentos de mecânica clássica previamente conhecidos. Tem-se:



$$\vec{r}(t + \Delta t) = \vec{r}(t) + \frac{d\vec{r}}{dt}\Delta t + \frac{\vec{v}^2}{d^2t \cdot 2!}\Delta t^2 = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{\vec{F}(t)}{2m}\Delta t^2 \quad (7)$$

$$\vec{r}(t - \Delta t) = \vec{r}(t) - \frac{d\vec{r}}{dt}\Delta t + \frac{\vec{v}^2}{d^2t \cdot 2!}\Delta t^2 = \vec{r}(t) - \vec{v}(t)\Delta t + \frac{\vec{F}(t)}{2m}\Delta t^2 \quad (8)$$

Nas Eqs. 7 e 8, temos as posições e velocidades das partículas e a série foi truncada no termo da derivada segunda para exemplificar o método. Assim, podemos somar as duas equações de forma a tornar-se:

$$\vec{r}(t + \Delta t) + \vec{r}(t - \Delta t) = 2\vec{r}(t) + \frac{\vec{F}(t)}{m}\Delta t^2 \quad (9)$$

Dividindo todos os termos da Eq. 8 por  $\Delta t$  e realocando o  $\vec{r}(t)$  para que tenhamos uma similaridade com os termos da direita das Eqs. 4 e 5, tem-se:

$$\frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t} = \frac{\vec{r}(t) - \vec{r}(t - \Delta t)}{\Delta t} + \frac{\vec{F}(t)}{m}\Delta t \quad (10)$$

E por fim, por analogia às Eqs. 4 e 5, tem-se:

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\vec{F}(t)}{m}\Delta t \quad (11)$$

ou seja, teremos a velocidade no tempo futuro, a partir da velocidade no tempo anterior e da força atuante em alguma partícula de massa ( $m$ ) no tempo atual ( $t$ ).

A partir da Eq. 11, tem-se todos os termos necessários para o cálculo da trajetória das partículas utilizando o algoritmo *Leap-Frog*. O *Verlet*, que originou este último, não será discutido aqui visto que não foi utilizado nas simulações desta tese (LEVITT; MEIROVITCH; HUBER, 1983). Mais detalhes podem ser obtidos nas referências a seguir, assim como de outros algoritmos disponíveis na literatura como o *Beeman* ou o *Velocity Verlet* (SCHOFIELD, 1973; SWOPE et al., 1982). A vantagem do *Leap-Frog* em relação aos demais é a rapidez e a reprodutibilidade dos

resultados, além disso, sua escolha foi reforçada devido a compatibilidade e implementação existente no software utilizado para realização dos cálculos.

#### 4.1.2 Modelos de solvatação

Os métodos de solvatação nas simulações moleculares são divididos em duas grandes classes, os modelos explícitos e os implícitos. Nos modelos implícitos, também conhecida como solvatação contínua, o meio é simulado através de uma constante dielétrica a depender do solvente escolhido. Esse método é geralmente utilizado quando se deseja fazer grandes simulações com um custo computacional menor já que nenhuma molécula de solvente é adicionada (ANANDAKRISHNAN et al., 2015). Contudo, possui a desvantagem de não tratar a natureza corpuscular das interações soluto-solvente na primeira camada de solvatação, principalmente em canais iônicos e bicamadas lipídicas, pelo motivo já mencionado. Esse método foi preterido nesta tese já que se tinha a intenção de obter detalhes exatamente da interação e da água com as proteínas simuladas (ALLEN; TILDESLEY, 1989).

Nos modelos explícitos, moléculas do solvente são adicionadas ao sistema de modo a modelar a atuação do solvente nas suas diversas características, sejam elas de pressão e densidade, ou de cunho químico por interações eletrostáticas e polaridade. A diferença entre os modelos explícitos está nos resultados dos parâmetros obtidos e a forma com que são construídos. Alguns modelos de água, por exemplo, funcionam melhor na descrição de interações químicas, enquanto outros modelam melhor o comportamento físico do solvente (FRENKEL; SMIT, 2001).

Nos ateremos aos modelos de água, visto que esse é o tipo de solvente utilizado em simulações de sistemas proteicos. Dentre esses modelos, destaca-se o SPC (*Simple Point Charge*) que foi utilizado em três publicações oriundas no período de escrita desta tese, entretanto, mais de 40 modelos de água já foram publicados provenientes de parâmetros quânticos, experimentais ou híbrido entre eles. O SPC possui característica de ser rígido, planar e utilizar o sistema de três pontos, ou seja, tratar somente os três átomos existentes (BERENDSEN et al., 1981). Outros modelos descrevem-na com 4, 5 e até 6 pontos pelos quais modelam a presença de pares isolados e geometrias diferentes (GUILLLOT, 2002). Esse modelo possui

resultados reprodutíveis e com boa concordância com resultados experimentais (MARK; NILSSON, 2001).

#### 4.1.3 Algoritmos de restrição

Para que a simulação de sistemas envolvendo macromoléculas em solvente explícito sejam realizadas com maior eficiência é imprescindível a utilização de algoritmos de restrição. Essas ressalvas matemáticas são empregadas na descrição de graus de liberdade a fim de que seu cálculo, através de um potencial harmônico por exemplo, torne-se dispensável. Seu uso possibilita, principalmente, que sejam tratados corretamente componentes de energia potencial relacionados à frequências de estiramento de ligação demasiadamente elevadas, como as que envolvem os átomos hidrogenoides ou grupos muito leves. Dessa forma são alcançados bons custos computacionais com um tamanho de passo razoável.

Alguns algoritmos realizam essa função. O LINCS (*Linear Constraint Solver*) é utilizado para moléculas grandes e são frequentemente utilizados para restrição de ligações covalentes (HESS et al., 1997). Já o SETTLE, possui uma capacidade restritiva menor (apenas 3 constantes) e é empregado na restrição de modelos rígidos de água, como o SPC comentado anteriormente (MIYAMOTO; KOLLMAN, 1992). Contudo, outros exemplos podem ser encontrados na literatura, como o SHAKE e o SHAPE, entre outros que encontram-se nas referências a seguir (BARTH et al., 1995; RYCKAERT; CICCOTTI; BERENDSEN, 1977; TAO; WU; BROOKS, 2012). Nos capítulos seguintes foram utilizados o LINCS e o SETTLE para restrição das ligações covalentes das proteínas e do modelo de água SPC respectivamente.

#### 4.2 Campos de Força

Por tratar-se de um modelo de átomos e moléculas, os graus de liberdade e a conectividade de cada átomo em determinada molécula devem ser considerados para que a descrição seja completa. Logo, dois termos principais são evidentes: os ligados e os não-ligados. Juntos, esses dois fatores somados caracterizam o potencial efetivo de interação entre duas moléculas em um momento específico.

Esses termos configuram-se como a parte mais importante do modelo, e portanto, a confiabilidade da simulação depende de quão bom esses parâmetros são para descrever determinado sistema.

Existem diversos grupos de parâmetros já conhecidos na literatura e por meio deles vários tipos de moléculas são estudadas (LOPES; GUVENCH; MACKERELL, 2015). Esse conjunto de dados que definem os valores ótimos de propriedades ligadas e não-ligadas é conhecido como campo de força. Os campos de força geralmente são criados no intuito de modelar um grupo de moléculas, como proteínas, aminoácidos, ácidos nucleicos, lipídios e outras moléculas orgânicas (ALLEN; TILDESLEY, 1989; FRENKEL; SMIT, 2001). Por isso a vasta quantidade de opções disponíveis na literatura para os mais diversos fins. Entre os mais utilizados estão o AMBER, o GROMOS e o CHARMM em suas dezenas de versões que vão sendo ajustadas na medida em que os resultados vão sendo publicados<sup>1</sup> (HUANG et al., 2017; MAIER et al., 2015; OOSTENBRINK et al., 2004). Além disso, eles podem ser parametrizados, conforme a necessidade do usuário para moléculas ausentes, gerando milhares de extensões com parâmetros calculados por pesquisadores do mundo inteiro.

Nessa tese, as simulações foram feitas utilizando o GROMOS96. Logo, os fatores que compõem a equação de potencial efetivo de interação serão detalhados em função das publicações originais desse campo de força. Esse potencial é o mesmo que foi apresentado na Eq. 1 deste capítulo. Em geral, diversos campos de força utilizam equações muito similares, contudo, alguns acrescentam particularidades que os tornam diferenciados para determinado grupo de moléculas.

O potencial utilizado pelo GROMOS96 (OOSTENBRINK et al., 2004) está descrito na equação abaixo e é dependente da distância e de parâmetros descritos nesse campo de força:

$$V_{\text{Total}} = V_{\text{Ligados}} + V_{\text{Não-Ligados}} \quad (12)$$

onde  $V_{\text{Ligados}}$  é o termo que descreve os parâmetros ligados enquanto que o  $V_{\text{Não-Ligados}}$  correspondente aos parâmetros não-ligados. Os termos ligados

---

<sup>1</sup> As referências citadas de cada um dos campos de força correspondem as versões ff14SB do AMBER, a 36m para o CHARMM e a 53a6 para o GROMOS96.

designam parâmetros de equilíbrio para ligação entre dois átomos, ângulo entre três átomos e o ângulo diedro formado entre semiplanos resultantes de dois pares de átomos. Já o  $V_{\text{Não-Ligados}}$  engloba as interações eletrostáticas e de van der Waals de todos os pares de átomos em uma distância previamente designada. Em suma, a Eq. 12 pode ser reescrita na forma expandida tornando-se:

$$V_{\text{Total}} = V_{\text{Est}} + V_{\text{Ang}} + V_{\text{Diedral}} + V_{\text{Coulomb}} + V_{\text{vdW}} \quad (13)$$

Cada um desses termos será brevemente discutido nas subseções a seguir.

#### 4.2.1 Estiramento de ligação

As ligações entre os pares de átomos são descritas por um potencial modelado pela Lei de Hooke na forma quadrática para um sistema de massas separadas por molas como exemplificado na Fig. 4.1a. O termo  $V_{\text{Est}}$ , mostrado na equação X abaixo, é um somatório de todas as ligações covalentes ( $N_r$ ) existentes no sistema entre grupos  $i$  e  $j$ :

$$V_{\text{Est}}(r_{ij}) = \frac{1}{2} \sum_{n=1}^{N_r} K_{r_n} [r_{ijn}^2 - r_{0n}^2]^2 \quad (14)$$

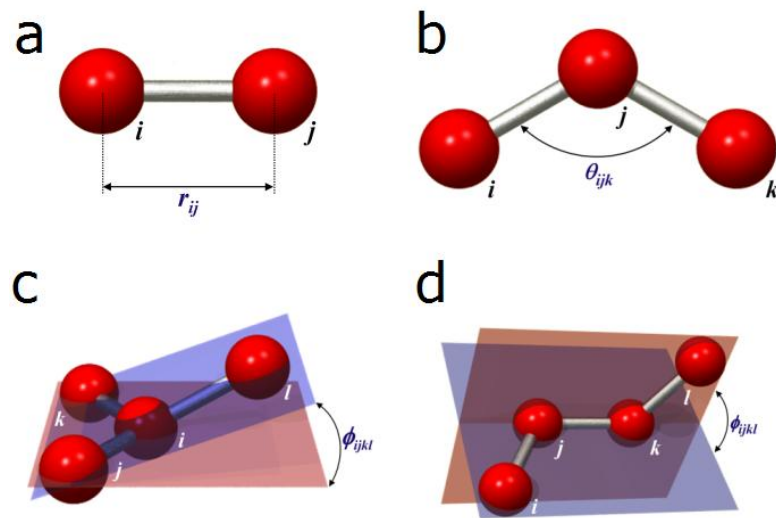
onde  $K_{r_n}$  é a constante de força da ligação  $n$ ,  $r_{ij}$  é a distância entre os grupos  $i$  e  $j$  em dado momento e  $r_0$  é a distância ideal entre eles.

#### 4.2.2 Deformação angular

De forma análoga ao  $V_{\text{Est}}$ , o potencial de deformação angular ( $V_{\text{Ang}}$ ) também funciona através de uma constante de força e um parâmetro ótimo. Nesse caso, o ângulo formado entre três grupos consecutivamente ligados ( $i$ ,  $j$  e  $k$ ) será a variável, como representado na Figura 4.1b. O termo  $V_{\text{Ang}}$  mostrado na equação X também é um somatório, desta vez de todos os ângulos formados entre ligações ( $N_\theta$ ) no sistema:

$$V_{\text{Ang}}(\theta) = \frac{1}{2} \sum_{n=1}^{N_{\theta}} K_{\theta_n} [\cos \theta_n - \cos \theta_{0_n}]^2 \quad (15)$$

onde  $K_{\theta_n}$  é a constante de força da  $n$ -ésima vibração angular,  $\theta_n$  é o ângulo associado a essa deformação e  $\theta_{0_n}$  é o seu valor ideal.



**Figura 4.1** – Esquema representativo dos termos ligados, ou seja, a) estiramentos de ligação, b) deformação angular, c) variação diedral imprópria e d) variação diedral própria utilizado pelo campo de força GROMOS96.

#### 4.2.3 Variações diedrais

O potencial diedral ( $V_{\text{Diedral}}$ ) pode ser dividido em dois tipos visto que essa angulação é diferente a depender da disposição espacial dos grupos envolvidos. Na Fig. 4.1 podemos ver a diferença entre esses dois tipos. No primeiro caso, Fig. 4.1c, conhecido como diedro impróprio, o ângulo resultante ( $\phi_{ijkl}$ ) ocorre entre os planos formados pelos grupos  $jkl$  e  $kij$ . Já no diedro próprio, Fig. 4.1d, não existe um grupo central que liga-se a três outros grupos como no diedro impróprio, sendo assim, o ângulo diedro resultante ( $\phi_{ijkl}$ ) é formado pelos planos  $ijk$  e  $jkl$ . Devido a essa diferença, a representação matemática de cada uma é feita independentemente e de forma diferenciada.

Para os diedros impróprios, o potencial segue a seguinte descrição:

$$V_{D. \text{ Impróprio}}(\phi) = \frac{1}{2} \sum_{n=1}^{N_\phi} K_{\phi_n} [\phi_n - \phi_{0_n}]^2 \quad (16)$$

onde  $K_{\phi_n}$  é a constante de força da  $n$ -ésima variação diedral,  $\phi_n$  é o ângulo associado a essa deformação entre os grupos  $ijkl$  e  $\phi_{0_n}$  é o seu valor ideal. Para o cálculo do ângulo  $\phi_{ijkl}$ , o *software* utiliza conhecimentos de álgebra linear. Inicialmente calcula-se o vetor normal ao plano formado pelos grupos  $jk$  ( $\vec{n}_{jkl}$ ) através do produto vetorial  $jk$  ( $\vec{v}_{jk}$ ) e  $kl$  ( $\vec{v}_{kl}$ ). Da mesma forma para o plano formado pelos grupos  $kij$  ( $\vec{n}_{kij}$ ) pelo produto dos vetores  $ki$  ( $\vec{v}_{ki}$ ) e  $ij$  ( $\vec{v}_{ij}$ ). Temos:

$$\vec{n}_{jkl} = \vec{v}_{jk} \times \vec{v}_{kl} \quad (17)$$

$$\vec{n}_{kij} = \vec{v}_{ki} \times \vec{v}_{ij} \quad (18)$$

Aplicando conceitos de álgebra linear, podemos encontrar o cosseno do ângulo  $\phi_{ijkl}$  e seu respectivo valor através da operação inversa,

$$\cos \phi_{ijkl} = \frac{\vec{n}_{jkl} \cdot \vec{n}_{kij}}{r_{n_{jkl}} r_{n_{kij}}} \quad (19)$$

$$\phi_{ijkl} = \arccos \left( \frac{\vec{n}_{jkl} \cdot \vec{n}_{kij}}{r_{n_{jkl}} r_{n_{kij}}} \right) \quad (20)$$

onde  $r_{n_{jkl}}$  e  $r_{n_{kij}}$  são os comprimentos dos respectivos vetores  $\vec{n}_{jkl}$  e  $\vec{n}_{kij}$ .

Para os diedros próprios, o potencial leva em consideração não só o somatório sobre todas as contribuições dos ângulos  $\phi_{ijkl}$  e a constante dessa variação, mas também a multiplicidade  $m_n$  e o deslocamento de fase  $\delta_n$ . A equação que o descreve é:

$$V_{D. \text{ Próprio}}(\phi) = \sum_{n=1}^{N_\phi} K_{\phi_n} [1 + \cos(\delta_n) \cos(m_n \phi_n)] \quad (21)$$

onde  $K_{\phi_n}$  é a constante de força da  $n$ -ésima variação diedral e  $\phi_n$  é o ângulo associado a essa deformação entre os grupos  $ijkl$ .

O deslocamento de fase ( $\delta_n$ ) tem valores restritos de 0 e  $\pi$ , e por tratar-se de uma cossenóide, só pode assumir valores de 1 e -1, funcionando na expressão como um indicativo para qual lado está sendo tratada a variação. Já o fator  $m_n$  indica a quantidade de mínimos existentes quando a ligação é rotacionada 360°. Logo, os valores possíveis são diferentes de zero, inteiros e positivos. A Eq. 21 pode ser obtida trabalhando-se as Eqs. 17 e 18 como feito para os diedros impróprios, contudo, com as conectividades correspondentes.

#### 4.2.4 Interações não-ligadas de van der Waals

As interações de curto alcance entre regiões em um determinado sistema são demasiadamente importantes para o sucesso de uma simulação clássica. Essas interações não-ligadas são calculadas como um somatório de todos os possíveis pares de grupos que se encontram dentro de determinado raio de corte através da função potencial de Lennard-Jones. A equação abaixo descreve esse potencial:

$$V_{\text{vdW}}(r_{ij}) = \sum_{\text{pares } i,j} \frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} \quad (22)$$

onde  $r_{ij}$  é a distância entre os átomos  $i$  e  $j$ ,  $C_{12}$  e  $C_6$  são respectivamente os coeficientes de repulsão e atração entre os pares  $i$  e  $j$ .

Os termos  $C_{12}$  e  $C_6$  dependem exclusivamente dos grupos envolvidos na interação, por isso, possuem valores tabelados no conjunto de parâmetros do campo de força que os usa ao identificar tal situação.

#### 4.2.5 Interações eletrostáticas não-ligadas

Juntamente com as interações de van der Waals, as interações entre cargas pontuais completam o potencial das interações não-ligadas em uma simulação clássica. Desde que os grupos, sejam eles átomos ou conjunto deles, possuem cargas parciais, elas interagem entre si modulando não só o comportamento de uma



molécula, como também sua interação com o meio, seja ele um solvente tratado explicitamente ou uma constante dielétrica utilizada para modelar um solvente implícito.

O potencial eletrostático que descreve as interações é dado pela lei de Coulomb:

$$V_{\text{Ele}}(r_{ij}) = \sum_{\text{pares } i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \quad (23)$$

onde  $q_i$  e  $q_j$  são as cargas  $i$  e  $j$  em questão,  $\epsilon_0$  e  $\epsilon_1$  são as permissividades do vácuo e a permissividade relativa do meio no qual se encontram os átomos e  $r_{ij}$  é a distância entre as cargas. Por padrão, o valor do  $\epsilon_1$  é igual a 1 e por isso que muitas vezes ele não é encontrado na literatura da área.

#### 4.2.6 Raio de corte e condições periódicas de contorno

Como visto em seções anteriores, o tratamento das cargas é função da distância entre os grupos interagentes. Esse conceito, mesmo que simplório, traz consequências importantes no planejamento e execução dos cálculos. Além disso, efeitos de superfície deveriam afetar consideravelmente um sistema tão pequeno como a caixa simulada e trazer implicações quando os resultados fossem comparados com dados experimentais macroscópicos. Devido a isso, os algoritmos atuais consideram dois artifícios para a interação dos grupos de moléculas simuladas: o raio de corte, para interações eletrostáticas não-ligadas (apresentadas na seção anterior), e as condições periódicas de contorno que anulam os efeitos de superfície.

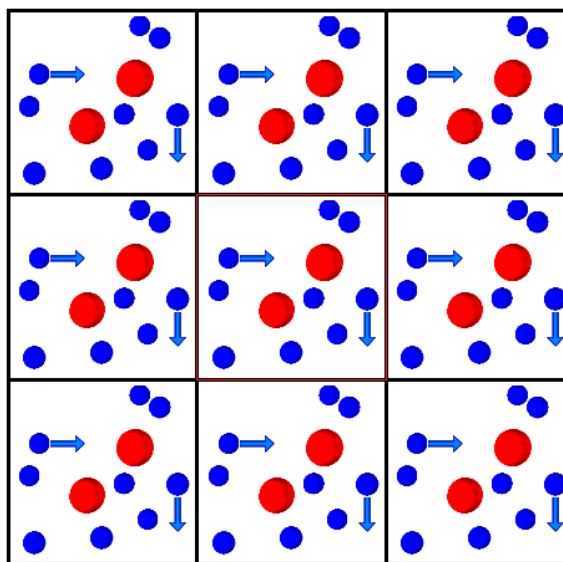
Já que a Eq. 23 considera um somatório de todos os pares possíveis de carga, em um sistema com milhares de moléculas de solvente e uma estrutura proteica enorme, o tempo necessário para computar todas essas influências seria inimaginável. Além disso, muitos desses pares de carga teriam contribuições ínfimas para o valor final visto que estariam à uma distância muito grande. Para resolver esse entrave, é necessário que seja imposto um limite de distância para busca dessas interações. Essa região, arbitrariamente escolhida, deve seguir alguns

critérios como a capacidade de *hardware* disponível, o tamanho do sistema e a finalidade de se realizar tal simulação (FRENKEL; SMIT, 2001). Dentro do raio de corte, como também é conhecida essa região, as interações entre as cargas são tratadas explicitamente, em contrapartida, tudo que circunda essa área também deve receber algum tipo de tratamento para que o sistema seja bem descrito (ALLEN; TILDESLEY, 1989).

Apesar de terem magnitude menor, essas interações que acontecem além do raio de corte ocorrem em grande quantidade e acabam tendo um peso considerável. Quando somadas, todas as entradas e saídas de moléculas nessa região, resultam em uma descontinuidade da energia do sistema, ou seja, elas não se cancelam e ocasionam uma má conservação da energia durante um ciclo de simulação. Para melhorar a modelagem do sistema, métodos foram desenvolvidos para tratar justamente essas contribuições e tornar o sistema contínuo além dessa região. Dois deles se destacam pelos resultados e são mais comumente usados.

O primeiro deles (utilizado nos Cap. 5 e 6) baseia-se na soma de Ewald (ALLEN; TILDESLEY, 1989). Esse método inicialmente foi proposto por Paul Peter Ewald para calcular energias de cristais iônicos e hoje é o mais usado para descrever as interações de cargas em longo alcance na química computacional (ALLEN; TILDESLEY, 1989; FRENKEL; SMIT, 2001). O método utiliza imagens repetidas do sistema presente dentro do raio de corte e através de transformadas de Fourier conseguem atenuar rapidamente esse potencial em uma distância consideravelmente maior. Derivações desse método como o PME (*Particle Mesh Ewald*) e o SPME (*Smooth Particle Mesh Ewald*) também são comumente usados (DARDEN; YORK; PEDERSEN, 1993; ESSMANN et al., 1995). Outra opção à esse problema no truncamento do raio de corte é o campo de reação (RF, *Reaction Field*) (TIRONI et al., 1995). Nesse caso, as cargas presentes dentro do raio de corte, chamado de cavidade nesse modelo, polarizam o meio externo, representado por uma constante dielétrica ( $\epsilon_{RF}$ ) que por sua vez atuará da mesma forma influenciando as moléculas dentro da cavidade na medida que a simulação acontece (ALLEN; TILDESLEY, 1989). Dessa forma, não há uma variação abrupta de energia visto que haverá sempre uma compensação entre as cargas da cavidade e o meio externo ( $\epsilon_{RF}$ ).

Mesmo com essa correção, a quantidade de moléculas presentes próximas à superfície da caixa simulada seria elevada frente ao número total das mesmas e a descrição de suas interações seria distorcida dos valores reais. Então, a aproximação de condições periódicas de contorno (do inglês, *Periodic Boundary Condition*, PBC) são empregadas nas simulações atuais. A técnica consiste em replicar nos eixos  $x$ ,  $y$  e  $z$  cópias idênticas do sistema real (caixa com todos os átomos a serem tratados) e está ilustrada na Fig. 4.2. Fazendo essa extrapolação ao infinito, elimina-se a superfície e os problemas a ela associados (ALLEN; TILDESLEY, 1989; FRENKEL; SMIT, 2001).



**Figura 4.2** – Esquema representativo da replicação da caixa central de simulação em dois eixos devido a aplicação das PBCs.

Ademais, para que os cálculos de distância entre partículas (utilizado em vários algoritmos dentro do programa como no cálculo de potenciais e forças), sejam corretamente calculadas, se faz obrigatório o uso de outra técnica chamada *Minimum Image Convention* (sem tradução direta), a qual considera a distância correta entre a partícula  $i$  e  $j$  como sendo a menor entre as opções de distância real entre o par ou distância entre a partícula  $i$  e a imagem replicada de  $j$  (ALLEN; TILDESLEY, 1989; FRENKEL; SMIT, 2001).

As PBCs irão impactar diretamente no tratamento das cargas e consequentemente na dinâmica das partículas durante a simulação e por isso, aliada ao raio de corte, são de extrema importância nesse tipo de cálculo para fins

de velocidade dos cálculos e reprodução correta dos dados experimentais desejados.

#### 4.3 Ferramentas de Análise

As propriedades observáveis em uma DM são obtidas através de cálculos baseados na movimentação e nas interações das partículas do sistema em função do tempo. A trajetória atômica (arquivo que contém todas as coordenadas e velocidades de todas as partículas do sistema em função do tempo) contém toda a informação necessária para realizarmos as análises necessárias.

Existem dezenas de ferramentas que são utilizadas para esse fim, assim como uma grande quantidade de softwares disponíveis, sejam eles comercialmente ou gratuitamente distribuídos. Outra opção encontrada pelos usuários dessa técnica baseia-se na construção de seu próprio algoritmo de análise que devido sua singularidade pode atender necessidades específicas de cada sistema simulado. Nesta tese somente softwares livres foram utilizados, além de um pequeno *script* escrito para realizar uma tarefa em específico. Os softwares utilizados foram o VMD (*Visual Molecular Dynamics*, versão 1.9.2) e o GROMACS (versões 4.6 e 5.0).

Este último também foi o escolhido para realizar todo o cálculo de DM apresentado na seção anterior. O GROMACS (*GRoningen MACHine for Chemical Simulation*) é uma das primeiras escolhas quando se pensa em cálculos de simulação molecular. É um pacote de vários programas que possui mais de dois milhões de linhas escritas em seu código-fonte. É um software bastante citado na literatura. Desde sua primeira publicação em 1995, seus artigos (de versões diferentes) já possuem juntos mais de 15 mil citações (ABRAHAM et al., 2015; BERENDSEN; VAN DER SPOEL; VAN DRUNEN, 1995; HESS et al., 2008; LINDAHL; HESS; VAN DER SPOEL, 2001; PRONK et al., 2013; VAN DER SPOEL et al., 2005). Outras opções como o AMBER (CASE et al., 2012), NAMD (PHILLIPS et al., 2005) e LAMMPS (PLIMPTON, 1995) também poderiam ser escolhidos mas além da gratuidade, a facilidade de manipulação, velocidade dos cálculos, compatibilidade de extensões, variedade de opções de campos de força e a abundância de ferramentas de análises fizeram com que o GROMACS fosse o selecionado.

#### 4.3.1 Raiz do desvio quadrático médio (do inglês, Root Mean Square Deviation, RMSD)

A estabilidade das estruturas geradas pela simulação pode ser avaliada pelo cálculo do RMSD. A análise baseia-se na escolha de uma estrutura de referência qualquer (geralmente a estrutura inicial,  $t=0.0\text{ns}$ ) e então o desvio é calculado para cada estrutura gerada pela simulação. Quanto mais distantes estiverem as estruturas analisadas, maior será o valor do desvio. A equação utilizada pelo programa *g\_rms* do GROMACS é:

$$\text{RMSD}(t) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[ \left( s_{i_x} - w_{i_x}(t) \right)^2 + \left( s_{i_y} - w_{i_y}(t) \right)^2 + \left( s_{i_z} - w_{i_z}(t) \right)^2 \right]} \quad (24)$$

onde  $n$  é o número de átomos,  $s_{i_x}$  é a posição do átomo  $i$  de referência no eixo  $x$  (e assim respectivamente para os eixos  $y$  e  $z$ ),  $w_{i_x}(t)$  é a posição do átomo  $i$  no eixo  $x$  no tempo  $t$  (e assim respectivamente para os eixos  $y$  e  $z$ ).

Em geral, são escolhidos somente alguns átomos do sistema para servir como referência, evitando que o  $n$  não seja demasiadamente grande e diminua a eficiência da medida. Para sistemas proteicos, o mais comum é que átomos de carbono alfa ou os átomos do *backbone* sejam utilizados para esse fim.

#### 4.3.2 Raiz da flutuação quadrática média (do inglês, Root Mean Square Fluctuation, RMSF)

A ideia de se calcular o RMSF surgiu a partir de resultados de RMSD. Trata-se da medida de variação da posição de determinada partícula, ou grupo delas, em relação a sua referência em determinado intervalo de tempo. Apesar de parecer com o RMSD, seus valores não necessariamente são correlacionados. Para explicar melhor, pode-se usar o exemplo de uma cadeia lateral de aminoácido numa simulação com a proteína. Caso as estruturas ao longo da simulação sejam diferentes da estrutura de partida, tem-se um alto valor de RMSD, porém se estas

estruturas foram semelhantes, não haverá flutuação estrutural, logo, o RMSF é quase nulo.

Dados de RMSF são utilizados nas DM's a fim de estimar a flexibilidade de determinado grupos, geralmente aminoácidos. Um alto valor de RMSF indica que houve grande oscilação na estrutura desses átomos, logo, pode-se afirmar que aquele grupo apresenta alta flexibilidade.

#### 4.3.3 Avaliação da distância

As distâncias entre grupos, sejam eles átomos, moléculas, fitas  $\beta$  ou  $\alpha$ -hélices, podem ser medidos e plotados em função do tempo. Nesta tese, essa distância foi calculada através do software VMD. A equação utilizada para tal é:

$$D_{i,j}(t) = |r_i(t) - r_j(t)| \quad (25)$$

onde  $D_{i,j}(t)$  é a distância entre os átomos  $i$  e  $j$  no tempo  $t$  e  $r_i(t)$  e  $r_j(t)$  são as coordenadas atômicas (tridimensionais) desses átomos num tempo específico.

O cálculo da distância durante o período de tempo simulado é importante porque pode demonstrar a aproximação e/ou afastamento de grupos, os quais podem ser relacionados desde a quebra/formação de ligações de hidrogênio até a estabilidade da estrutura secundária e terciária de proteínas. Nesta tese por exemplo, o movimento de abertura e fechamento de xilanasas, descrito no Cap. 2 pode ser verificado pelo cálculo das distâncias entre aminoácidos específicos.

#### 4.3.4 Energia Potencial de interação (do inglês, *Interaction Potential Energy, IPE*)

A energia potencial de interação (IPE) pode dar uma medida quantitativa da interação não-ligada entre dois grupos. A equação que descreve esse parâmetro baseia-se no potencial descrito acima nas Eqs. 22 e 23. Para um cálculo da energia de interação entre duas moléculas  $i$  e  $j$ , tem-se:

$$\text{IPE}_{i,j} = \sum_i^{N_i} \sum_{j \neq i}^{N_j} V_{ij} \quad (26)$$

onde  $N_i$  e  $N_j$  são todos os átomos constituintes das moléculas  $i$  e  $j$ , e o  $V_{ij}$  é o potencial de interação entre os átomos dessas moléculas.

Nesta tese, os Cap. 5 e 6 utilizaram essa ferramenta para descrição da interação entre grupos, sejam eles intramoleculares, como a interação de aminoácidos entre dois domínios proteicos, ou intermoleculares, no cálculo da interação entre moléculas do solvente e cadeias laterais de aminoácidos da xilanase (SILVA et al., 2017). Como mostrado nas Eq. 22 e 23, esse cálculo de energia é dependente de uma distância que deve ser especificada pelo usuário. Durante os capítulos, na medida em que for empregado esse parâmetro, a distância do raio de corte também será apresentada já que variará para cada caso.

#### 4.3.5 Quantidade de Resíduos Interfaciais

Uma análise realizada ao longo de uma simulação DM da quantidade de resíduos em uma determinada distância da interface entre domínios proteicos torna possível a obtenção de informações úteis sobre aquele complexo de interesse. Uma abordagem semelhante para indicar o grau de afinidade da interface proteína-proteína poderia ser também feita através da energia potencial de interação, mas por tratar-se de uma estrutura muito grande a ser considerada no cálculo, dados de pontes salinas muito interagentes energeticamente ou resíduos com alta repulsão poderia mascarar o somatório final. Além disso, a velocidade com que o resultado é gerado (mesmo para grandes sistemas) e a não obrigatoriedade de tratamento dos arquivos de entrada trazem uma grande versatilidade do *script* em comparação, por exemplo, ao sistema de *rerun* da trajetória amplamente utilizado pelo GROMACS no qual há necessidade da criação de arquivos de adicionais empregando outras rotinas.

Nesse intuito, um pequeno *script* escrito em *Python* foi escrito em parceria com o aluno Fernando Rodrigues Maciel para o cálculo da quantidade de resíduos em determinada distância a partir de um conjunto de arquivos tipo *pdb* (*Protein Data Bank files*). Esses arquivos podem ser oriundos de uma trajetória de simulação

molecular ou estruturas únicas provenientes de bancos de dados e saídas de outros programas de *docking* ou modelagem por homologia. Há também a possibilidade de escolha por parte do usuário de qual será o raio de corte considerado a partir da interface entre os domínios, garantindo a flexibilidade necessária para tratar diferentes sistemas. Nesta tese, uma distância de 4,5 Å foi usada para definir a região interfacial.

#### 4.3.6 *Alanine Scanning Computacional*

A influência de determinado aminoácido na estabilização ou desestabilização de uma interface proteína-proteína pode ser mensurada através do *alanine scanning* computacional. Essa técnica tem origens experimentais e é bastante utilizada na engenharia de proteínas para avaliar os principais pontos de interação em determinada interface após a mutação do resíduo de interesse por um resíduo de alanina, cuja cadeia lateral (constituída por um grupo metil) é não-volumosa e não-polar. Dessa forma, a substituição serve como indicativo para avaliar interações eletrostáticas principalmente, e para alguns casos, a influência estrutural do aminoácido mutante na afinidade entre dois domínios (KORTEMME; BAKER, 2002).

A abordagem computacional baseia-se no mesmo princípio. Inicialmente é calculada uma função de energia livre de Gibbs para cada domínio isolado e seus resultados são subtraídos do valor obtido para o complexo entre eles. Então, faz-se a mutação do resíduo de interesse por alanina e o cálculo é refeito. A diferença entre os dois estados dá um indicativo de quão importante é aquele resíduo para a interface. Caso apresente valores de  $\Delta\Delta G_{\text{LIGAÇÃO}} > 1$  kcal/mol, os resíduos são considerados como *hotspots*, ou seja, resíduos com grande importância energética para estabilização da região estudada (DELANO, 2002).

Vários servidores *on-line* e *softwares* livres calculam essa propriedade com grande rapidez mesmo para interfaces extensas. Nesta tese, foi utilizado o *ROBETTA program* (KORTEMME; KIM; BAKER, 2004) devido sua relação de tempo utilizado para realização do cálculo com os valores compatíveis com dados experimentais.

Nesta tese, o *alanine scanning* computacional foi utilizado para avaliar a influência de aminoácidos na interface proteína-proteína formada entre uma xilanase



produzida pela bactéria *Bacillus subtilis* (XynA) e uma proteína receptora de xilose (XBP) conforme detalhado no Cap. 7.

#### 4.3.7 Área Superficial Acessível ao Solvente (ASAS)

A análise da área superficial acessível ao solvente tem a função de calcular a exposição de biomoléculas, ou sub regiões dela, às moléculas do solventes que as circundam. Os valores obtidos através da análise de ASAS geralmente são dados em Å<sup>2</sup> ou nm<sup>2</sup> e variam de acordo com o tamanho da biomolécula, geralmente proteínas e peptídeos já enovelados.

A forma clássica do cálculo dessa propriedade é feita através de um algoritmo que simula uma esfera de raio previamente definido rolando sobre uma superfície da molécula ou região de interesse (SHRAKE; RUPLEY, 1973). Contudo, outras abordagens foram desenvolvidas com o passar do tempo variando seu custo computacional e, conseqüentemente, sua confiabilidade nos resultados obtidos (HAYRYAN et al., 2005; KLENIN et al., 2011; WEISER; SHENKIN; STILL, 1999). Nesta tese um subprograma presente no GROMACS foi utilizado para o cálculo das ASAS por resíduo durante o tempo simulado. O método implementado no programa foi desenvolvido por F. Eisenhaber e colaboradores (EISENHABER et al., 1995).

Essa esfera utilizada em geral tem o raio parecido com uma molécula de água ou do solvente desejado. Caso consiga alcançar uma maior região no espaço através da movimentação dessa esfera imaginária, aquela região também estará disponível para encontro com o solvente durante uma simulação. Por exemplo, é comum que cadeias laterais de proteínas sejam mais expostas e tenham elevados valores de ASAS enquanto que os átomos do *backbone* apresentam valores praticamente nulos.

## Referências - Capítulo 4

- ABRAHAM, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, set. 2015.
- ALLEN, M. P.; TILDESLEY, D. J. **Computer simulation of liquids**. New York, NY, USA: Clarendon Press, 1989.
- ANANDAKRISHNAN, R. et al. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. **Biophysical Journal**, v. 108, n. 5, p. 1153–1164, mar. 2015.
- ANASTAS, P. T.; KIRCHHOFF, M. M. Origins, Current Status, and Future Challenges of Green Chemistry †. **Accounts of Chemical Research**, v. 35, n. 9, p. 686–694, set. 2002.
- ARGAMAN, N.; MAKOV, G. Density functional theory: An introduction. **American Journal of Physics**, v. 68, n. 1, p. 69, 2000.
- BARNOUD, J.; MONTICELLI, L. Coarse-Grained Force Fields for Molecular Simulations. In: **Kukol A. (eds) Molecular Modeling of Proteins. Methods in Molecular Biology (Methods and Protocols)**, vol 1215. Humana Press, New York, NY
- BARTH, E. et al. Algorithms for constrained molecular dynamics. **Journal of Computational Chemistry**, v. 16, n. 10, p. 1192–1209, out. 1995.
- BERENDSEN, H. J. C. et al. Interaction Models for Water in Relation to Protein Hydration. In: **Intermolecular Forces**. Dordrecht: Pullman B., 1981. p. 331–342.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1–3, p. 43–56, 2 set. 1995.
- BOND, P. J.; HOLYOAKE, J.; IVETAC, A.; KHALID, S.; SANSOM, M. S. P. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. **Journal of Structural Biology**, v. 157, n. 3, p. 593–605, mar. 2007.
- CASE, D. A. et al. **AMBER 12** University of California, San Francisco, , 2012. Disponível em: <citeulike-article-id:10779586>
- CHUNG, L. W. et al. The ONIOM Method and Its Applications. **Chemical Reviews**, v. 115, n. 12, p. 5678–5796, jun. 2015.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method

for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089, 1993.

DE MORAIS, C. DE L. M. et al. Integrating a Smartphone and Molecular Modeling for Determining the Binding Constant and Stoichiometry Ratio of the Iron(II)–Phenanthroline Complex: An Activity for Analytical and Physical Chemistry Laboratories. **Journal of Chemical Education**, v. 93, n. 10, p. 1760–1765, 11 out. 2016.

DELANO, W. L. Unraveling hot spots in binding interfaces: progress and challenges. **Current Opinion in Structural Biology**, v. 12, n. 1, p. 14–20, fev. 2002.

EISENHABER, F. et al. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. **Journal of Computational Chemistry**, v. 16, n. 3, p. 273–284, mar. 1995.

ESSMANN, U. et al. A smooth particle mesh Ewald method. **Journal of Chemical Physics**, v. 103, n. 19, p. 8577–8593, 1995.

FIELD, M. J.; BASH, P. A.; KARPLUS, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. **Journal of Computational Chemistry**, v. 11, n. 6, p. 700–733, jul. 1990.

FIRME, C. L.; MONTEIRO, N. K. V.; SILVA, S. R. B. QTAIM and NCI analysis of intermolecular interactions in steroid ligands binding a cytochrome P450 enzyme – Beyond the most obvious interactions. **Computational and Theoretical Chemistry**, v. 1111, p. 40–49, jul. 2017.

FRENKEL, D.; SMIT, B. **Understanding Molecular Simulation**. 2nd. ed. Orlando, FL, USA: Academic Press, Inc., 2001.

GELPI, J. et al. Molecular dynamics simulations: advances and applications. **Advances and Applications in Bioinformatics and Chemistry**, p. 37, nov. 2015.

GUILLOT, B. A reappraisal of what we have learnt during three decades of computer simulations on water. **Journal of Molecular Liquids**, v. 101, n. 1–3, p. 219–260, nov. 2002.

HADLEY, K. R.; MCCABE, C. Coarse-grained molecular models of water: a review. **Molecular Simulation**, v. 38, n. 8–9, p. 671–681, jul. 2012.

HAYRYAN, S. et al. A new analytical method for computing solvent-accessible surface area of macromolecules and its gradients. **Journal of computational**

**chemistry**, v. 26, n. 4, p. 334–43, mar. 2005.

HESS, B. et al. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, set. 1997.

HESS, B. et al. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 3, p. 435–447, mar. 2008.

HUANG, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. **Nature Methods**, v. 14, n. 1, p. 71–73, 7 jan. 2017.

KARPLUS, M.; KURIYAN, J. Molecular dynamics and protein function. **Proceedings of the National Academy of Sciences**, v. 102, n. 19, p. 6679–6685, 10 maio 2005.

KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. **Nature Structural Biology**, v. 9, n. 9, p. 646–652, set. 2002.

KLEINJUNG, J.; FRATERNALI, F. Design and application of implicit solvent models in biomolecular simulations. **Current Opinion in Structural Biology**, v. 25, p. 126–134, abr. 2014.

KLENIN, K. V. et al. Derivatives of molecular surface area and volume: Simple and exact analytical formulas. **Journal of Computational Chemistry**, v. 32, n. 12, p. 2647–2653, set. 2011.

KORTEMME, T.; BAKER, D. A simple physical model for binding energy hot spots in protein-protein complexes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 22, p. 14116–14121, 2002.

KORTEMME, T.; KIM, D. E.; BAKER, D. Computational alanine scanning of protein-protein interfaces. **Science's STKE : signal transduction knowledge environment**, v. 2004, n. 219, p. pl2, 2004.

LEVINE, I. N. **Quantum Chemistry**. 6th. ed. New York: Pearson Prentice Hall, 2009.

LEVITT, M.; MEIROVITCH, H.; HUBER, R. Integrating the equations of motion. **Journal of Molecular Biology**, v. 168, n. 3, p. 617–620, ago. 1983.

LINDAHL, E.; HESS, B.; VAN DER SPOEL, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. **Journal of Molecular Modeling**, v. 7, n. 8, p. 306–317, ago. 2001.

LOPES, P. E. M.; GUVENCH, O.; MACKERELL, A. D. Current Status of Protein Force Fields for Molecular Dynamics Simulations. In: [s.l: s.n.]. p. 47–71.

MAIER, J. A. et al. ff14SB: Improving the Accuracy of Protein Side Chain and

Backbone Parameters from ff99SB. **Journal of Chemical Theory and Computation**, v. 11, n. 8, p. 3696–3713, 11 ago. 2015.

MARK, P.; NILSSON, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. **The Journal of Physical Chemistry A**, v. 105, n. 43, p. 9954–9960, nov. 2001.

MIYAMOTO, S.; KOLLMAN, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. **Journal of Computational Chemistry**, v. 13, n. 8, p. 952–962, out. 1992.

OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1656–1676, out. 2004.

PHILLIPS, J. C. et al. Scalable molecular dynamics with NAMD. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1781–1802, dez. 2005.

PLIMPTON, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. **Journal of Computational Physics**, v. 117, n. 1, p. 1–19, mar. 1995.

PRONK, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics**, v. 29, n. 7, p. 845–854, 1 abr. 2013.

RIBEIRO, L. F. et al. A xylose-stimulated xylanase-xylose binding protein chimera created by random nonhomologous recombination. **Biotechnology for biofuels**, v. 9, p. 119, 2016.

ROUX, B.; SIMONSON, T. Implicit solvent models. **Biophysical Chemistry**, v. 78, n. 1–2, p. 1–20, abr. 1999.

RYCKAERT, J.-P.; CICCOTTI, G.; BERENDSEN, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **Journal of Computational Physics**, v. 23, n. 3, p. 327–341, mar. 1977.

SCHOFIELD, P. Computer simulation studies of the liquid state. **Computer Physics Communications**, v. 5, n. 1, p. 17–23, jan. 1973.

SHAW, D. E. et al. Atomic-level characterization of the structural dynamics of proteins. **Science (New York, N.Y.)**, v. 330, n. 6002, p. 341–6, 15 out. 2010.

SHRAKE, A.; RUPLEY, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. **Journal of molecular biology**, v. 79, n. 2, p. 351–71, 15 set. 1973.

SILVA, S. B. et al. The role of local residue environmental changes in thermostable

mutants of the GH11 xylanase from *Bacillus subtilis*. **International Journal of Biological Macromolecules**, v. 97, p. 574–584, abr. 2017.

SLATER, J. C. A Simplification of the Hartree-Fock Method. **Physical Review**, v. 81, n. 3, p. 385–390, 1 fev. 1951.

SWOPE, W. C. et al. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. **The Journal of Chemical Physics**, v. 76, n. 1, p. 637–649, jan. 1982.

SZABO, A.; OSTLUND, N. S. **Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory**. New York, NY, USA: Dover, 1996.

TAO, P.; WU, X.; BROOKS, B. R. Maintain rigid structures in Verlet based Cartesian molecular dynamics simulations. **The Journal of Chemical Physics**, v. 137, n. 13, p. 134110, 7 out. 2012.

TIRONI, I. G. et al. A generalized reaction field method for molecular dynamics simulations. **The Journal of Chemical Physics**, v. 102, n. 13, p. 5451–5459, abr. 1995.

VAN DER SPOEL, D. et al. GROMACS: fast, flexible, and free. **Journal of computational chemistry**, v. 26, n. 16, p. 1701–18, dez. 2005.

VAN GUNSTEREN, W. F.; BERENDSEN, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. **Molecular Simulation**, v. 1, n. 3, p. 173–185, mar. 1988.

WARSHEL, A.; LEVITT, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. **Journal of Molecular Biology**, v. 103, n. 2, p. 227–49, 15 maio 1976.

WEISER, J.; SHENKIN, P. S.; STILL, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). **Journal of Computational Chemistry**, v. 20, n. 2, p. 217–230, 30 jan. 1999.

YOUNG, D. C. **Computational chemistry: a practical guide for applying techniques to real-world problems**. New York ; Chichester: Wiley-Interscience, 2001.

---

## Capítulo 5

### **O efeito de mutações locais em uma xilanase GH11 termoestável: uma análise por dinâmica molecular <sup>2</sup>**

Como visto previamente no Cap. 3, as xilanases têm uma grande importância biotecnológica, mais notadamente nas áreas de branqueamento de papel, atuando na redução da quantidade de produtos químicos à base de cloro utilizados nos processos, e na produção de etanol lignocelulósico, já que possibilita uma maior eficiência reacional e menor dano ambiental quando comparada aos agentes inorgânicos fortes classicamente utilizados (PAËS; BERRIN; BEAUGRAND, 2012). No entanto, os processos industriais nos quais essas enzimas são empregadas utilizam condições severas de temperatura e pH, e com isso, surge uma limitação associada à sua utilização nesses meios. Esse entrave ocorre devido ao processo de desnaturação total ou parcial que afeta os domínios catalíticos, e consequentemente, a eficiência da catálise enzimática que são drasticamente diminuídas ou anuladas.

Para tanto, torna-se imprescindível a criação de enzimas estáveis e eficientes em condições adversas de pH e/ou sob altas temperaturas. Essa estratégia pode ser aplicada com auxílio de ferramentas teóricas e experimentais. O advento das técnicas de biologia molecular permitiu não só a produção em larga escala de enzimas modificadas, mas também sua modificação por mutagênese sítio-dirigida ou evolução programada (PORTER; RUSLI; OLLIS, 2016). A engenharia de proteínas combina a análise da estrutura da proteína 3D com a biologia molecular para evoluir ou desenvolver novas enzimas com propriedades catalíticas ou termodinâmicas

---

<sup>2</sup> Conteúdo desse capítulo foi publicado. Ver referência : (SILVA et al., 2017)

melhoradas (PAËS; BERRIN; BEAUGRAND, 2012). Já as técnicas de bio e quimioinformática, como simulações de dinâmica molecular (DM), comparação de sequências de aminoácidos e análises de estruturas 3D de proteínas têm sido amplamente utilizadas para promover a compreensão das relações estrutura/função da proteína. Devido à correlação entre as propriedades funcionais, estruturais, energéticas e dinâmicas das proteínas, as simulações de dinâmica molecular podem fornecer informações úteis para programas de engenharia enzimática (KARPLUS; KURIYAN, 2005). Uma comparação entre os resultados de simulações DM de estruturas proteicas permite descrever processos relacionados à catálise e desnaturação térmica. Isso aumenta a possibilidade de identificar regiões que possam promover a termoestabilidade sem afetar a atividade (MURAKAMI et al., 2005) e identificar mudanças estruturais que gerem termoestabilidade através de mudanças nas propriedades físico-químicas locais resultantes de mutações direcionadas.

Não foi estabelecida uma teoria unificadora em relação à base estrutural da termoestabilidade proteica, uma vez que o aumento da estabilidade é resultante de uma conjunção de múltiplos fatores. No caso das ferredoxinas e da proteína ribossômica termofílica L30e, a termoestabilidade foi atribuída à contribuições eletrostáticas por redes de pares de íons (CHAN; YU; WONG, 2011; PERUTZ; RAIDT, 1975). Isso está de acordo com estudos de simulações DM que envolvem laminarinases demonstrando que a termoestabilidade da proteína é dependente do número de pares de íons em posições de superfície específicas (BLEICHER et al., 2011), otimização de ligações de hidrogênio (HB) e interações eletrostáticas (TORREZ; SCHULTEHENRICH; LIVESAY, 2003).

Vários estudos mostraram que as proteínas termofílicas exibem flexibilidade reduzida nas temperaturas ambiente em relação aos seus homólogos mesofílicos (JAENICKE, 1996; VIEILLE; GREGORYZEIKUS, 1996; ZHANG et al., 2016). Por outro lado, as observações experimentais tanto da beta amilase (FITTER; HEBERLE, 2000) como da esterase de *Bacillus acidocaldarius* (D'AURIA et al., 2000), combinado a simulações DM de subtilisina-E (COLOMBO; MERZ, 1999) e rubredoxina (GROTTESE et al., 2002) mostraram que as proteínas dos organismos termofílicos são mais flexíveis do que seus homólogos mesofílicos. Esta aparente discrepância talvez seja resolvida pela observação de que, embora as enzimas



termoestáveis geralmente apresentam uma flexibilidade geral reduzida para manter a estabilidade, eles mantêm a flexibilidade local da cadeia polipeptídica em regiões funcionalmente importantes (RADESTOCK; GOHLKE, 2011). Este ponto de vista pode explicar por que a modulação da flexibilidade estrutural das proteínas é uma estratégia comum para aumentar a termoestabilidade (FITTER; HEBERLE, 2000; GIVER et al., 1998; RATHI et al., 2016; VAN DEN BURG et al., 1998; VIEILLE; GREGORYZEIKUS, 1996). Além disso, estudos da catálise enzimática das xilanases GH11 demonstraram que os movimentos internos das proteínas podem estar correlacionados com a função deles (MURAKAMI et al., 2005; TÖRRÖNEN; HARKKI; ROUVINEN, 1994).

O empacotamento dos resíduos no núcleo hidrofóbico também pode influenciar a termoestabilidade, como demonstrado no caso da aldeído ferredoxina oxidoreductase (CHAN et al., 1995). Além disso, um grupo de mutações pontuais que criam cavidades apresentaram uma maior termoestabilidade para a lisozima T4 (MATTHEWS, 1995). Outros fatores que são conhecidos por aumentar a termoestabilidade da proteína são: redes melhoradas de ligação de hidrogênio (SHIH; KIRSCH, 1995), estabilização de alfa-hélices (FACCHIANO; COLONNA; RAGONE, 1998), encurtamento de dobras entre elementos estruturais secundários (HARDY et al., 1994), otimização de superfícies eletrostáticas (TORREZ; SCHULTEHENRICH; LIVESAY, 2003), e aumento da solvatação em regiões específicas da proteína (BLEICHER et al., 2011). Os esforços para melhorar a termoestabilidade enzimática por engenharia de proteínas concentraram-se, portanto, na melhoria das interações intra e intermoleculares por introdução de pontes dissulfeto (WATANABE; FUKADA; ISHIKAWA, 2016) e a introdução de resíduos aromáticos empilhados (GEORIS et al., 2008).

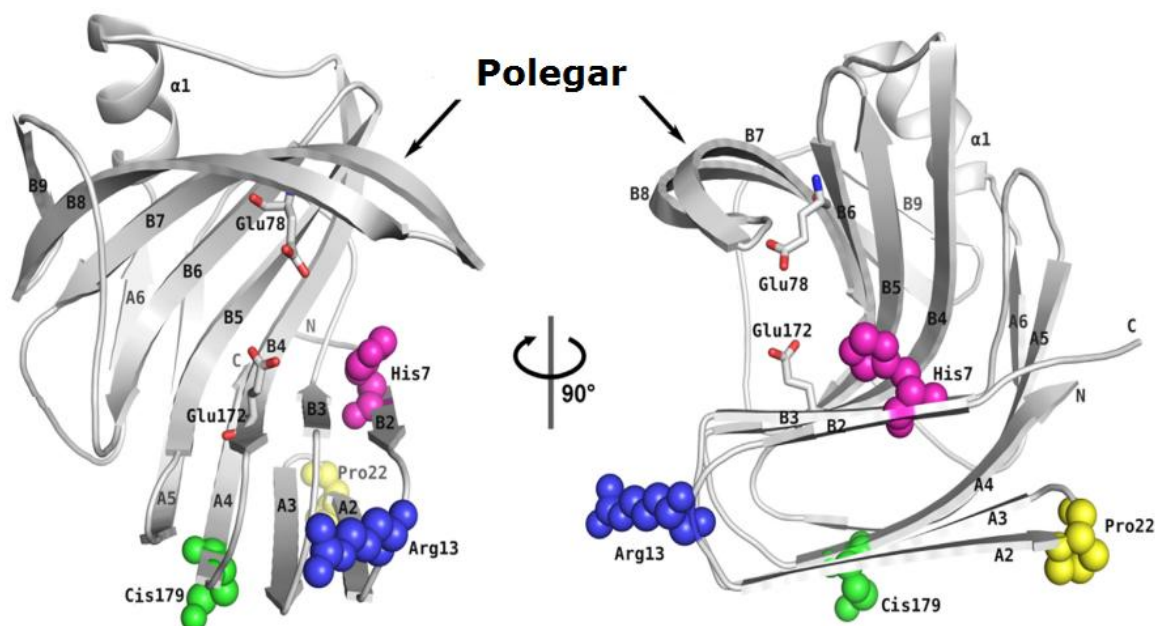
A evolução direta de proteínas é uma estratégia amplamente reconhecida para a engenharia enzimática e é particularmente útil quando a base estrutural de uma função desejada é desconhecida ou envolve múltiplos determinantes estruturais. Em um estudo anterior (RULLER et al., 2008), técnicas experimentais de evolução de proteínas foram utilizadas para gerar mutantes termoestáveis da xilanase A mesofílica GH11 de *Bacillus subtilis* (XynA\_WT). O mutante XynA\_WT mais termoestável obtido por evolução dirigida continha as quatro mutações Q7H, G13R, S22P e S179C (doravante referido como XynA-4x) e demonstrou um

aumento de 20°C na temperatura catalítica ótima ( $T_{OPT}$ ) em relação a enzima XynA\_WT ( $T_{OPT} = 55^{\circ}\text{C}$ ).

A fim de explicar a termoestabilidade da XynA-4x em termos estruturais dessas quatro mutações, simulações DM foram feitas nas  $T_{OPT}$  da XynA\_WT (55°C) e XynA-4x (75°C). Os resultados obtidos aliados aos dados experimentais provenientes de cristalografia de raios-x e análises termodinâmicas (parceria com grupos de pesquisa da USP) reforçam o argumento de que a caracterização de mudanças estruturais locais resultantes de mutações individuais contribui para a compreensão dos múltiplos fatores que influenciam a termoestabilidade das xilanases de GH11.

### 5.1 Detalhes Computacionais

As simulações foram feitas usando o campo de força GROMOS 53A6 (OOSTENBRINK et al., 2004) através do pacote GROMACS 5.0 (PRONK et al., 2013). Os cálculos foram feitos fixando a temperatura (T), número de átomos (N) e volume (V). O pH dos sistemas foi ajustado para 7,0 através da análise dos pKas das cadeias laterais de cada aminoácido constituinte das enzimas realizado via sub-programas do pacote GROMACS. As estruturas cristalinas de partida da XynA\_WT e da XynA-4x foram obtidas através do banco de dados de proteínas (*Protein Data Bank*, PDB) com os códigos 1XXN e 5K9Y respectivamente. Quatro mutações foram feitas para criar essa enzima melhorada por meio da substituição manual das cadeias laterais dos aminoácidos de interesse. A glutamina na posição 7 foi modificada por uma histidina (Q7H), a glicina na posição 13 por uma arginina (G13R), a serina na posição 22 por uma prolina (S22P) e a serina da posição 179 por uma cisteína (S179C). A representação 3D da enzima e as posições das mutações estão evidenciadas na Fig. 5.1. Duas temperaturas distintas foram escolhidas, justamente as  $T_{OPT}$  de cada enzima. No caso, 328 K para a XynA\_WT e 348 K para XynA-4x. Os tempos totais de simulação variaram entre 80 ns e 100 ns de acordo com o tempo de convergência de cada proteína. As simulações para ambos os sistemas foram feitas sucessivamente, ou seja, a estrutura de partida para a simulação em 348 K da XynA\_WT e XynA-4x foram as últimas geradas na respectiva simulação em 328 K a fim de que seja modelada uma escada de aquecimentos.



**Figura 5.1** - Representação 3D da enzima XynA-4x em dois ângulos distintos. Elementos da estrutura secundária foram definidos de acordo com as xilanases *T. reesei* I e II (GUIMARÃES et al., 2009). Os dois resíduos catalíticos (E78 e E172) estão evidenciados em tubos cinzas. Os resíduos modificados H7, R13, P22 e C179 estão representados em esferas roxas, azuis, amarelas e verdes, respectivamente.

A temperatura foi controlada usando o algoritmo V-Rescale (BUSSI; DONADIO; PARRINELLO, 2007). As ligações covalentes envolvendo os átomos de hidrogênio e a geometria das moléculas de água, tipo SPC, foram mantidas através dos algoritmos LINCS e SETTLE (seção 4.1.3 desta tese) respectivamente (HESS et al., 1997; MIYAMOTO; KOLLMAN, 1992). As equações de movimento nas quais são baseadas as simulações clássicas foram integradas usando o algoritmo Leap-Frog (seção 4.1.1 desta tese) com o passo temporal de 2,0 fs (VAN GUNSTEREN; BERENDSEN, 1988). As velocidades iniciais foram ajustadas usando a distribuição de Maxwell-Boltzmann. As interações de longo alcance foram modeladas usando o método PME (ver seção 4.2.6) com um corte de 1,5 nm (DARDEN; YORK; PEDERSEN, 1993). Esse raio de corte foi o mesmo utilizando para as interações de van der Waals. As simulações foram feitas dentro de sistemas dodecaédricos completos com água e íons (cloretos e de sódio) adequados para garantir a neutralidade do meio. Cada sistema foi submetido a minimizações de energia através do algoritmo *steepest descent* (ADCOCK; MCCAMMON, 2006) e períodos

curtos de simulação (200 ps) para alcançar um equilíbrio inicial sem energias altamente repulsivas de átomos muito próximos. Na fase de equilíbrio, a proteína foi mantida fixa enquanto o solvente e os íons ficaram livres para se mover, a fim de evitar energias intermoleculares altamente repulsivas na configuração inicial do experimento.

As análises de desvio quadrático médio (RMSD, seção 4.3.1), que avaliam a estabilidade ao longo do tempo das estruturas geradas em relação ao  $t=0$  ns, e de flutuação média quadrada (RMSF, seção 4.3.2) foram calculadas através das posições dos carbonos alfa ( $C\alpha$ ) dos *backbones* proteicos. As ligações de hidrogênio intramoleculares ( $HB_{INTRA}$ ) e intermoleculares ( $HB_{INTER}$ ) foram identificadas por dois critérios, *i*) a distância (seção 4.3.3) entre o átomo doador e receptor da densidade eletrônica, que deveria ser menor que 0,35 nm, e *ii*) o ângulo doador-hidrogênio-receptor que deveria ser menor que  $30^\circ$  para que satisfizesse a condição de existência dessa interação (ARUNAN et al., 2011; VAN DER SPOEL et al., 2006). A energia potencial de interação, doravante IPE também foi computada (seção 4.3.4) para resíduos e aminoácidos em sua vizinhança ou moléculas de água em torno do resíduo modificado. Além disso, a análise da flexibilidade local foi avaliada por valores de flutuação média quadrada por resíduo (RMSF, ver seção 4.3.2).

## 5.2 Resultados e Discussão

Cada posição modificada na enzima XynA-4x junto com os resíduos vizinhos foram analisados e comparados com a XynA\_WT para elucidar os fatores que contribuem para o aumento da termoestabilidade.

### *5.2.1 Mutação Q7H*

As interações intramoleculares com os resíduos da vizinhança dentro de 6,0 Å (incluindo Y5, W6, N8, W9, V37, G39, Y69, Y166 e V168) foram analisados nas temperaturas de 328 K e 348 K para ambas XynA\_WT e XynA-4x. Além disso, a primeira camada de solvatação que compreende as moléculas de água dentro de 3,5 Å a partir da posição 7 também foram incluídas na análise energética. Em 328 K os valores de IPE foram -0,3 kcal/mol (XynA\_WT) e -18,9 kcal/mol (XynA-4x), e em

348 K foram -9,9 kcal/mol para XynA\_WT e -19,1 kcal/mol para a XynA-4x. Isso sugere que a energia potencial de interação média da região que cerca a H7 na XynA-4x é essencialmente constante em ambas as temperaturas, mas é mais forte quando comparada a mesma região da XynA\_WT.

A HB<sub>INTRA</sub> específica entre Q7 (XynA\_WT) e H7 (XynA-4x) com os resíduos da vizinhança também foi analisada. Embora a HB<sub>INTRA</sub> entre o oxigênio carbonílico da posição 7 e o nitrogênio da amida da cadeia principal da G39 esteja conservado em ambas as temperaturas, a HB<sub>INTRA</sub> Q7-G39 na XynA\_WT esteve presente em 99% do tempo simulado a 328K e 89% a 348K. A HB<sub>INTRA</sub> H7-G39 na XynA-4x esteve presente em 94% do tempo simulado em ambas as temperaturas. A HB<sub>INTRA</sub> conservada na XynA\_WT envolvendo a cadeia lateral do resíduo na posição 7 e a Y5, esteve presente em 48% e 38% do tempo simulado para Q7-Y5 a 328 K e 348 K, respectivamente, enquanto que para H7-Y5 a ocorrência foi de 38% a 328 K e 51% a 348 K.

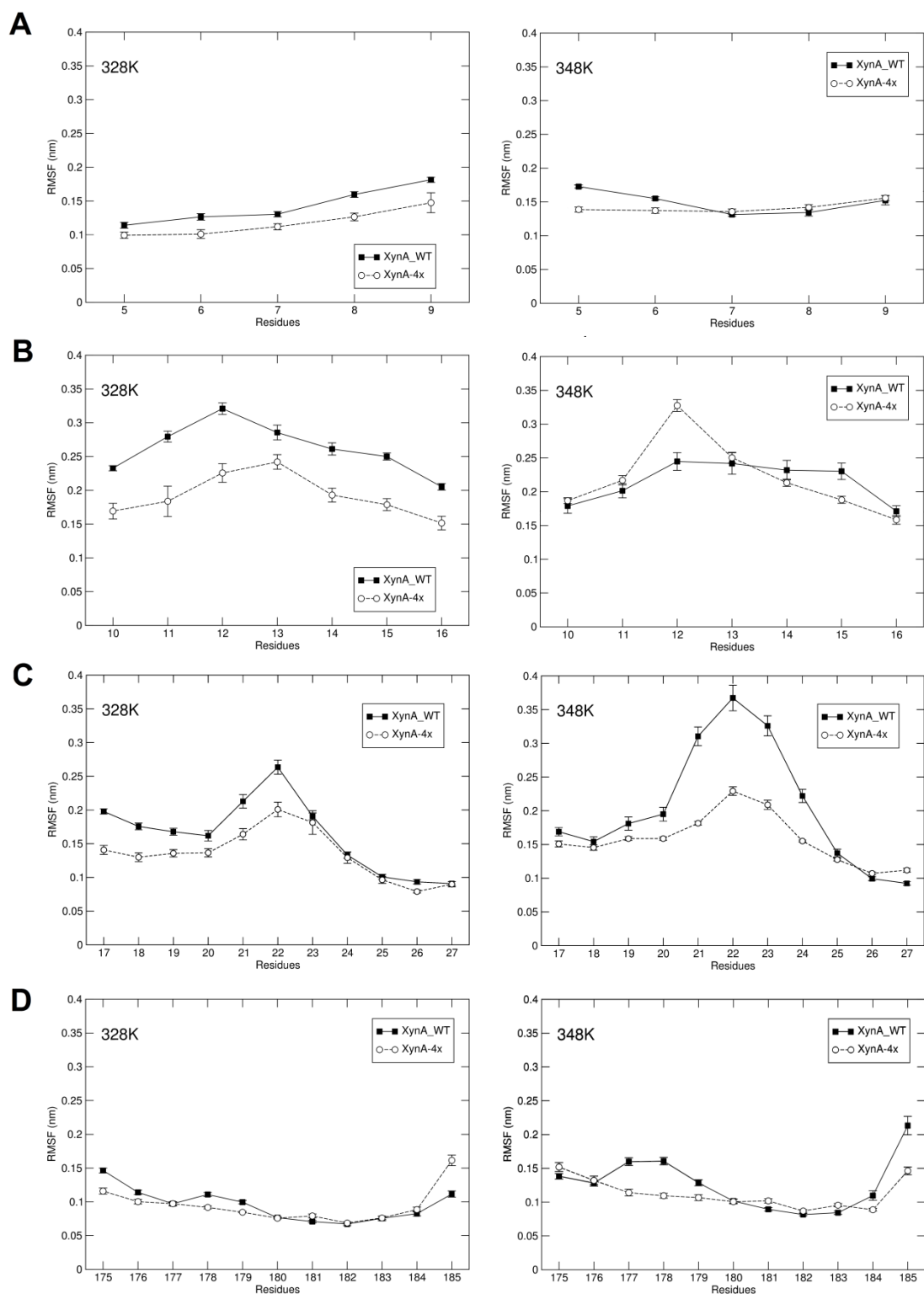
As análises de flexibilidade local são feitas através da flutuação quadrática média (RMSF) por resíduo (Fig. 5.2A). Para essa mutação, o RMSF mostra que não existe diferença nas posições 5–7 entre a XynA-4x e a XynA\_WT em 348 K. Em 328 K (a T<sub>OPT</sub> da XynA\_WT), uma pequena redução nos valores de RMSF nessa região (posição 5 a 9) é observada, sugerindo que os resíduos vizinhos são menos flexíveis com a H7 (na XynA-4x) em comparação com a Q7 (na XynA\_WT).

A análise comparativa das simulações XynA\_WT e XynA-4x revelaram que a mutação Q7H favorece ligações de hidrogênio com aminoácidos responsáveis pela maior estabilização da fita- $\beta$  N-terminal. Os átomos da cadeia principal da H7 também interagem com a G39 na XynA-4x, e suas cadeias laterais fazem ligações de hidrogênio mediadas por água que estabilizam a fita B2. Esses contatos contribuem para uma energia de interação mais favorável com os átomos vizinhos e resultam numa configuração mais estável quando comparado à mesma região na XynA\_WT.

### 5.2.2 Mutação G13R

As dobras- $\beta$  que conectam as fitas- $\beta$  nas xilanases GH11 são caracterizadas pela grande quantidade de glicina, e simulações moleculares prévias demonstraram

que essas regiões são altamente flexíveis (MURAKAMI et al., 2005; VIEIRA; DEGRÈVE; WARD, 2009).



**Figura 5.2** - Flutuações quadráticas médias (RMSF) por resíduo da XynA\_WT (quadrado preto) e XynA-4x (círculo branco) calculadas em 328 K e 348 K para os átomos da cadeia principal das (A) posições 5–9; (B) posições 10–16; (C) posições 17–27 e (D) posições 175–185. Cada ponto na Fig.3

*apresenta a média dos valores de RMSF obtidos em janelas de 8 ns e as barras de erro indicam o erro padrão da média.*

A posição 13 está localizada na dobra que conecta as fitas- $\beta$  B2 e A2 (resíduos 12 ao 14). Além disso, a introdução de um resíduo carregado na superfície da proteína está envolvido na formação de novas ligações de hidrogênio com as cadeias laterais da N32 e Q175. Fig. 5.2B compara o RMSF da região dos resíduos 10 ao 16 na XynA-4x e a XynA\_WT a 328 K e 348 K. A 328 K na mutante XynA-4x, a Arg13 e os resíduos da vizinhança demonstram uma considerável diminuição da flexibilidade. A 348 K a rigidez total dessa região é similar exceto pelo aumento da flexibilidade da G12 na mutante XynA-4x. Além das mudanças de flexibilidade observadas, a mutação G13R também tem um efeito significativo nas interações com a vizinhança. O hidrogênio da cadeia lateral da G13 na XynA\_WT foi substituído por um grande grupo carregado positivamente na R13 da XynA-4x (Tabela 5.1). Consequentemente, essa posição torna-se bem solvatada e forma HB<sub>INTRA</sub> com os resíduos próximos (D11, G12, G14, G34 e F36).

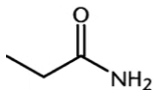
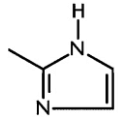
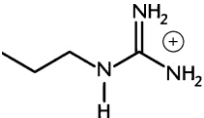
As simulações mostram que a mutação G13R favorece a energia potencial de interação média com os resíduos da vizinhança e com a primeira camada de solvatação. Essa mudança resultou em uma diminuição na energia de -13,2 kcal/mol na XynA\_WT à -28,9 kcal/mol na XynA-4x a 328K, e de -16,0 kcal/mol na XynA\_WT à -27,3 kcal/mol na XynA-4x a 348K.

Uma menor flexibilidade conformacional da R13 em XynA-4x também é observada na dobra que liga as fitas A2 e B2. Além disso, a cadeia lateral da R13 no XynA-4x pode formar ligações de hidrogênio com outros resíduos expostos ao solvente na superfície da proteína, o que contribui ainda mais para a estabilidade dessa dobra. A modificação na natureza da cadeia lateral na posição 13 resultou no aumento de interações, incluindo uma elevação expressiva na solvatação com energias significativamente mais atrativas no XynA-4x em comparação com XynA\_WT.

As mutações Q7H e G13R no XynA-4x aumentam o número de cadeias laterais carregadas positivamente na proteína. Resultados provenientes de análises termodinâmicas obtidos experimentalmente (SILVA et al., 2017) sugerem que as interações eletrostáticas resultantes da inserção desses resíduos contribuem para o aumento da termoestabilidade da enzima. Essa análise corrobora os resultados das

simulações DM e indica uma maior energia de interação das posições 7 e 13 com átomos vizinhos no XynA-4x, como apresentado na Tabela 5.1. A energia de interação calculada inclui contribuições de interações intra e intermoleculares, e é consistente com as contribuições entálpicas conhecidas de hidratação eletrostática (GILSON; HONIG, 1991) e quantidade maior de ligações de hidrogênio intramoleculares. Estes efeitos são particularmente evidentes para o mutante Q7H e podem explicar o aumento da entalpia observada experimentalmente ( $\Delta H_{\text{EXP}}$ ) em relação ao XynA\_WT em pH baixo (SILVA et al., 2017). As simulações indicam que o aumento das interações intramoleculares com os átomos da cadeia principal em torno das posições 7 e 13 resulta em um RMSF menor para a cadeia principal no XynA-4x. Além disso, as ligações de hidrogênio mediadas pelo solvente observadas na mutação Q7H imobilizam moléculas de solvente na superfície da proteína e estas interações de ligação de hidrogênio adicionais podem contribuir para a estabilidade da proteína. Um estudo recente com mutantes CutA1 hipertermoestáveis demonstrou que as interações carga-carga contribuem para a estabilização da proteína na faixa de temperatura 380-410 K. A solvatação de resíduos carregados abaixo de 380 K reduz a entropia do estado desnaturado, estabilizando assim a proteína (MATSUURA et al., 2015).

**Tabela 5.1** - Energia Potencial de Interação dos resíduos 7 e 13 com a vizinhança (raio de 6,0 Å) obtidos através das simulações DM em 328K e 348K para XynA\_WT e XynA-4x.

Mutação	Enzima	IPE (kcal/mol)		Cadeia Lateral
		328K	348K	
Q7H	XynA_WT	-0,3	-9,9	
	XynA-4x	-18,9	-19,1	
G13R	XynA_WT	-13,2	-16,0	H
	XynA-4x	-28,9	-27,3	



### 5.2.3 Mutação S22P

Os resíduos das posições 20 à 22 estão localizados na dobra que conecta as fitas- $\beta$  A2 e A3 (Fig. 5.1). Os resíduos N20, G21 e P22 adotam ângulos diedros mais favoráveis na XynA-4x se comparados com a estrutura da XynA\_WT, possivelmente com uma consequência da redução da flexibilidade conformacional do esqueleto imposta pela mutação em questão. Isso por sua vez resulta em um aumento da capacidade de fazer ligações HB<sub>INTRA</sub> na região. Além disso, a rigidez da dobra que conecta as fitas- $\beta$  A2 e A3 é aumentada pela inserção do resíduo de prolina. Isso pode contribuir também para a estabilidade da região C-terminal já que o átomo de hidrogênio da cadeia principal da G24 forma ligações de hidrogênio com o nitrogênio da cadeia principal da V184. Análises de RMSF demonstram que todos os resíduos que cercam a posição 22 apresentam uma drástica diminuição da flexibilidade como resultado da mutação S22P e esse efeito é particularmente evidente a 348 K (Fig. 5.4C).

A mutação na posição 22 na fita- $\beta$  A2 também contribui para a estabilidade da região N-terminal da proteína. A introdução do anel de pirrolidina da P22 na XynA-4x não só apresenta menos liberdade conformacional que a S22 na enzima XynA\_WT, mas também restringe o espaço conformacional dos resíduos vizinhos. Consequentemente, N20, G21 e P22 adotaram conformações mais favoráveis e contribuíram para as interações mais eficazes entre as fitas- $\beta$  A2, A3 e A4 na XynA-4x em comparação com a proteína XynA\_WT.

A introdução de um resíduo de prolina na posição 22 na XynA-4x também reduzirá a flexibilidade da cadeia polipeptídica no seu estado não enovelado e, portanto, é provável que diminua a entropia nesse estado. Isto reforça a observação experimental de que os efeitos entrópicos desempenham papel fundamental na termoestabilização da XynA-4x. Esses resultados foram publicados recentemente pelo grupo de pesquisa parceiro da USP através de mutantes que introduziam cargas na estrutura da XynA\_WT para demonstrar que a estabilidade térmica da XynA\_WT pode ser alcançada com várias estratégias de engenharia de proteínas através de diferentes parâmetros termodinâmicos (ALPONTI; FONSECA MALDONADO; WARD, 2016).

Nossos resultados sugerem que as mutações G7H, G13R e S22P conferem termoestabilidade ao aumentar a estabilidade da região das folhas- $\beta$  do domínio dos dedos formada pelas fitas A2 e B2, a dobra N-terminal e a fita- $\beta$  A4 da cadeia polipeptídica. A introdução de uma ponte de dissulfeto intramolecular entre o final das regiões N- e C-terminal da xilanase de GH11 de *Bacillus circulans* também provocou um aumento dramático na termoestabilidade (WATANABE; FUKADA; ISHIKAWA, 2016). Demonstrou-se ainda que a integridade do sítio ativo a temperaturas mais elevadas depende da estabilidade da fita- $\beta$  A1 (WATANABE; FUKADA; ISHIKAWA, 2016). Significativamente, os mutantes nas xilanases de GH11 mais termostáveis introduzem aminoácidos que estabilizam a região N-terminal (DUMON et al., 2008; MIYAZAKI et al., 2006; PALACKAL et al., 2004; ZHANG et al., 2010), o que indica que a região N-terminal de xilanases de GH11 é altamente suscetível à desnaturação térmica e influencia diretamente na estabilidade global da proteína.

#### 5.2.4 Mutação S179C

A posição 179 é um dos resíduos do início da fita- $\beta$  B4 e é exposta na superfície da proteína (Fig. 5.1). As análises de RMSF revelaram que embora não existam diferenças significativas na posição 179 entre a XynA-4x e a XynA\_WT, as posições 177 e 178 mostraram flexibilidade reduzida na XynA-4x a 348K (Fig. 5.2D).

Como ilustrado na Fig. 5.1, o meio imediato à posição 179 compreende os aminoácidos localizados nas fitas- $\beta$  A3 e A5, e os resíduos S179 na XynA\_WT e C179 na XynA-4x apresentam perfis alternativos de ligações de hidrogênio com os resíduos vizinhos. Embora os átomos da cadeia principal da S179 não formem HB<sub>INTRA</sub>, a simulação indica uma única HB<sub>INTER</sub> é formada com uma ocorrência de 57,1% e 13,8% a 328 K e 348 K, respectivamente. A cadeia lateral da S179 mostra um potencial significativo para formar HB<sub>INTER</sub> com o solvente. Na temperatura de 328 K, uma única HB<sub>INTER</sub> é detectada em 42,4% do tempo de simulação, enquanto que a HB<sub>INTER</sub> com duas moléculas de solvente foi observada em 44,1% do tempo. Ligações de hidrogênio intermoleculares com três moléculas de água também foram identificadas, mas somente em 5,6% do tempo. A 348 K a ocorrência de uma, duas e três HB<sub>INTER</sub> foi de 46,7%, 40,1% e 6,0%, respectivamente. Do contrário, a cadeia

principal da C179 não forma HB<sub>INTER</sub> com moléculas do solvente mas apresenta HB<sub>INTRA</sub> entre o oxigênio carbonílico e o nitrogênio amídico da A59 na vizinhança da fita- $\beta$  A5 com 87,0% e 78,9% de ocorrência a 328 K e 348 K, respectivamente.

Com base nas análises feitas, o mecanismo pelo qual a mutação S179C confere a termoestabilidade não é claro. A mutação S179C conserva o número de átomos da cadeia lateral e apresenta apenas um pequeno aumento no volume da cadeia lateral. A mutação S179C não resulta em alterações estruturais nesta região e a análise de flexibilidade de simulações DM não mostrou diferenças significativas entre XynA-4x e XynA\_WT.

No entanto, o resíduo de serina na XynA\_WT é um ponto de solvatação em uma região de superfície, de outra forma dominada por resíduos hidrofóbicos e a substituição de oxigênio pelo enxofre resulta em um aumento da hidrofobicidade nessa posição. A mutação S179C elimina a hidroxila de cadeia lateral da serina, resultando na perda de interações favoráveis de hidratação com solvente, mas pode-se esperar que eleve a entropia do solvente, estabilizando assim a XynA-4x. Embora a análise dos dados disponíveis para esta posição não seja conclusiva, vale salientar que a substituição da S179C em uma superfície hidrofóbica resultará na correspondência das propriedades físico-químicas da cadeia lateral com os resíduos vizinhos nessa região.

Essas propriedades termodinâmicas podem ser correlacionadas com fatores estruturais e dinâmicos para explicar o aumento da termoestabilidade do mutante XynA-4x. (i) Aumento da rigidez do backbone: a rigidez aumentada do backbone aparece como um fator fundamental para conferir termoestabilidade, em particular para as mutações nas posições 13 e 22. Nesses dois casos, o aumento da termoestabilidade ocorre nas regiões de maior flexibilidade na proteína, ou seja, as dobras- $\beta$  no domínio dos dedos. (ii) Solvação e estabilização local: a substituição de cadeias laterais pouco solvatadas por cadeias laterais muito solvatadas resulta no aumento de interações com moléculas de água e é a causa da contribuição para a termoestabilidade das mutações nas posições 7 e 13. Além disso, a otimização das interações locais com resíduos vizinhos promove conformações mais energeticamente favoráveis (iii) Contatos hidrofóbicos: a mutação Ser179Cys foi identificada como potencialização de contatos hidrofóbicos. Uma cadeia lateral altamente solvatada localizada perto de outros resíduos hidrofóbicos foi substituída

por um grupo não solvatado, e os efeitos na termoestabilidade das propriedades de cadeia lateral correspondentes nas posições de superfície requerem um estudo mais aprofundado.

### 5.3 Conclusão

Os resultados de simulação molecular em conjunto com os dados experimentais de cristalografia de raios-X e análises termodinâmicas do mutante quádruplo XynA-4x sugerem possíveis estratégias para a engenharia de xilanases termoestáveis da família GH11. As mutações em regiões que apresentam pouca flexibilidade deveriam preservar a rigidez, mas a substituição delas pode ser escolhida para afetar favoravelmente outras propriedades, como a solvatação ou interações hidrofóbicas. A mutação em uma região flexível da proteína, como as dobras- $\beta$ , deveriam visar o aumento da rigidez local, a fim de aumentar a termoestabilidade proteica. Além disso, uma vez que a relação interação/flexibilidade não é óbvia e altamente dependente da estrutura local da proteína, propomos que as interações intra e intermoleculares e a análise da flexibilidade do *backbone* devem ser unidas e realizadas região por região da proteína. O presente estudo não explica apenas o efeito de termoestabilidade de uma xilanase de GH11 observado em experimentos anteriores de evolução direta, mas também fornece informações úteis para o planejamento de outros mutantes GH11 termoestáveis por estratégias de projeto racionais.

## Referências - Capítulo 5

- ADCOCK, S. A.; MCCAMMON, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. **Chemical Reviews**, v. 106, n. 5, p. 1589–1615, maio 2006.
- ALPONTI, J. S.; FONSECA MALDONADO, R.; WARD, R. J. Thermostabilization of *Bacillus subtilis* GH11 xylanase by surface charge engineering. **International journal of biological macromolecules**, v. 87, p. 522–8, jun. 2016.
- ARUNAN, E. et al. Definition of the hydrogen bond (IUPAC Recommendations 2011). **Pure and Applied Chemistry**, v. 83, n. 8, p. 1637–1641, 8 jan. 2011.
- BLEICHER, L. et al. Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. **The journal of physical chemistry. B**, v. 115, n. 24, p. 7940–9, 23 jun. 2011.
- BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of chemical physics**, v. 126, n. 1, p. 014101, 7 jan. 2007.
- CHAN, C.-H.; YU, T.-H.; WONG, K.-B. Stabilizing Salt-Bridge Enhances Protein Thermostability by Reducing the Heat Capacity Change of Unfolding. **PLoS ONE**, v. 6, n. 6, p. e21624, 24 jun. 2011.
- CHAN, M. K. et al. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. **Science (New York, N.Y.)**, v. 267, n. 5203, p. 1463–9, 10 mar. 1995.
- COLOMBO, G.; MERZ, K. M. Stability and Activity of Mesophilic Subtilisin E and Its Thermophilic Homolog: Insights from Molecular Dynamics Simulations. **Journal of the American Chemical Society**, v. 121, n. 29, p. 6895–6903, jul. 1999.
- D'AURIA, S. et al. The esterase from the thermophilic eubacterium *Bacillus acidocaldarius*: structural-functional relationship and comparison with the esterase from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. **Proteins**, v. 40, n. 3, p. 473–81, 15 ago. 2000.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089, 1993.
- DUMON, C. et al. Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. **The Journal of biological chemistry**, v. 283,

n. 33, p. 22557–64, 15 ago. 2008.

FACCHIANO, A. M.; COLONNA, G.; RAGONE, R. Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. **Protein engineering**, v. 11, n. 9, p. 753–60, set. 1998.

FITTER, J.; HEBERLE, J. Structural equilibrium fluctuations in mesophilic and thermophilic alpha-amylase. **Biophysical journal**, v. 79, n. 3, p. 1629–36, set. 2000.

GEORIS, J. et al. An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: Structural basis and molecular study. **Protein Science**, v. 9, n. 3, p. 466–475, 31 dez. 2008.

GILSON, M. K.; HONIG, B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. **Journal of computer-aided molecular design**, v. 5, n. 1, p. 5–20, fev. 1991.

GIVER, L. et al. Directed evolution of a thermostable esterase. **Proceedings of the National Academy of Sciences**, v. 95, n. 22, p. 12809–12813, 27 out. 1998.

GROTTESE, A. et al. Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. **Proteins**, v. 46, n. 3, p. 287–94, 15 fev. 2002.

GUIMARÃES, B. G. et al. The MX2 macromolecular crystallography beamline: a wiggler X-ray source at the LNLS. **Journal of synchrotron radiation**, v. 16, n. Pt 1, p. 69–75, jan. 2009.

HARDY, F. et al. The effect of engineering surface loops on the thermal stability of *Bacillus subtilis* neutral protease. **Protein engineering**, v. 7, n. 3, p. 425–30, mar. 1994.

HESS, B. et al. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, set. 1997.

JAENICKE, R. Glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: Strategies of protein stabilization. **FEMS Microbiology Reviews**, v. 18, n. 2–3, p. 215–224, maio 1996.

KARPLUS, M.; KURIYAN, J. Molecular dynamics and protein function. **Proceedings of the National Academy of Sciences**, v. 102, n. 19, p. 6679–6685, 10 maio 2005.

MATSUURA, Y. et al. Thermodynamics of protein denaturation at temperatures over 100 °C: CutA1 mutant proteins substituted with hydrophobic and charged residues. **Scientific Reports**, v. 5, p. 15545, 26 out. 2015.

MATTHEWS, B. W. Studies on protein stability with T4 lysozyme. **Advances in**

**protein chemistry**, v. 46, p. 249–78, 1995.

MIYAMOTO, S.; KOLLMAN, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. **Journal of Computational Chemistry**, v. 13, n. 8, p. 952–962, out. 1992.

MIYAZAKI, K. et al. Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. **The Journal of biological chemistry**, v. 281, n. 15, p. 10236–42, 14 abr. 2006.

MURAKAMI, M. T. et al. Correlation of temperature induced conformation change with optimum catalytic activity in the recombinant G/11 xylanase A from Bacillus subtilis strain 168 (1A1). **FEBS Letters**, v. 579, n. 28, p. 6505–6510, 21 nov. 2005.

OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1656–1676, out. 2004.

PAËS, G.; BERRIN, J.-G.; BEAUGRAND, J. GH11 xylanases: Structure/function/properties relationships and applications. **Biotechnology advances**, v. 30, n. 3, p. 564–92, 2012.

PALACKAL, N. et al. An evolutionary route to xylanase process fitness. **Protein science : a publication of the Protein Society**, v. 13, n. 2, p. 494–503, fev. 2004.

PERUTZ, M. F.; RAIDT, H. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. **Nature**, v. 255, n. 5505, p. 256–259, 15 maio 1975.

PORTER, J. L.; RUSLI, R. A.; OLLIS, D. L. Directed Evolution of Enzymes for Industrial Biocatalysis. **ChemBioChem**, v. 17, n. 3, p. 197–203, fev. 2016.

PRONK, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics**, v. 29, n. 7, p. 845–854, 1 abr. 2013.

RADESTOCK, S.; GOHLKE, H. Protein rigidity and thermophilic adaptation. **Proteins**, v. 79, n. 4, p. 1089–108, abr. 2011.

RATHI, P. C. et al. Application of Rigidity Theory to the Thermostabilization of Lipase A from Bacillus subtilis. **PLOS Computational Biology**, v. 12, n. 3, p. e1004754, 22 mar. 2016.

RULLER, R. et al. Thermostable variants of the recombinant xylanase A from Bacillus subtilis produced by directed evolution show reduced heat capacity changes. **Proteins**, v. 70, n. 4, p. 1280–93, mar. 2008.

SHIH, P.; KIRSCH, J. F. Design and structural analysis of an engineered thermostable chicken lysozyme. **Protein Science**, v. 4, n. 10, p. 2063–2072, out. 1995.

SILVA, S. B. et al. The role of local residue environmental changes in thermostable mutants of the GH11 xylanase from *Bacillus subtilis*. **International Journal of Biological Macromolecules**, v. 97, p. 574–584, abr. 2017.

TORREZ, M.; SCHULTEHENRICH, M.; LIVESAY, D. R. Conferring Thermostability to Mesophilic Proteins through Optimized Electrostatic Surfaces. **Biophysical Journal**, v. 85, n. 5, p. 2845–2853, nov. 2003.

TÖRRÖNEN, A.; HARKKI, A.; ROUVINEN, J. Three-dimensional structure of endo-1,4-beta-xylanase II from *Trichoderma reesei*: two conformational states in the active site. **The EMBO Journal**, v. 13, n. 11, p. 2493–2501, 1 jun. 1994.

VAN DEN BURG, B. et al. Engineering an enzyme to resist boiling. **Proceedings of the National Academy of Sciences**, v. 95, n. 5, p. 2056–2060, 3 mar. 1998.

VAN DER SPOEL, D. et al. Thermodynamics of Hydrogen Bonding in Hydrophilic and Hydrophobic Media. **The Journal of Physical Chemistry B**, v. 110, n. 9, p. 4393–4398, mar. 2006.

VAN GUNSTEREN, W. F.; BERENDSEN, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. **Molecular Simulation**, v. 1, n. 3, p. 173–185, mar. 1988.

VIEILLE, C.; GREGORYZEIKUS, J. Thermozyms: Identifying molecular determinants of protein structural and functional stability. **Trends in Biotechnology**, v. 14, n. 6, p. 183–190, jun. 1996.

VIEIRA, D. S.; DEGRÈVE, L.; WARD, R. J. Characterization of temperature dependent and substrate-binding cleft movements in *Bacillus circulans* family 11 xylanase: a molecular dynamics investigation. **Biochimica et biophysica acta**, v. 1790, n. 10, p. 1301–6, out. 2009.

WATANABE, M.; FUKADA, H.; ISHIKAWA, K. Construction of Thermophilic Xylanase and Its Structural Analysis. **Biochemistry**, v. 55, n. 31, p. 4399–4409, 9 ago. 2016.

ZHANG, S. et al. Five mutations in N-terminus confer thermostability on mesophilic xylanase. **Biochemical and biophysical research communications**, v. 395, n. 2, p. 200–6, 30 abr. 2010.

ZHANG, X.-F. et al. A general and efficient strategy for generating the stable enzymes. **Scientific Reports**, v. 6, n. 1, p. 33797, 26 dez. 2016.



---

## Capítulo 6

### **Análise via dinâmica molecular da estabilidade e movimentação de uma quimera xilanase-XBP <sup>3</sup>**

Assim como no capítulo anterior, o objetivo das simulações de dinâmica molecular continua sendo desenvolvimento biotecnológico de processos que utilizam a enzima xilanase. Sua aplicabilidade e flexibilidade já foram anteriormente discutidas, bem como seu potencial para uma maior utilização em outros nichos produtivos, como a produção do etanol lignocelulósico. Da mesma forma que as mutações pontuais podem garantir propriedades otimizadas de pH e temperatura, outros métodos podem (e devem) ser empregados a fim de conferir uma alta eficiência reacional em processos industriais nos quais a xilanase atua como catalisador.

A catálise enzimática que ocorre na sacarificação do material lignocelulósico foi discutida no Cap. 3, e, em síntese, trata-se da quebra do substrato de xilana em xiloses. Esse processo feito por xilanases e outras hidrolases é geralmente realizado com uma alta concentração de sólidos para minimizar o uso da água em processos industriais, e essas condições inevitavelmente resultam no acúmulo de altas concentrações dos produtos. Esses produtos de reação frequentemente inibem as enzimas que são usadas no processo de sacarificação (ANDRIĆ et al., 2010; JÖNSSON; ALRIKSSON; NILVEBRANT, 2013; TEUGJAS; VÄLJAMÄE, 2013). Portanto, superar a inibição causada pelo produto reacional é um pré-requisito para o uso prático de enzimas em bioprocessos, além de servir como uma abordagem

---

<sup>3</sup> Parte do conteúdo desse capítulo foi publicado. Ver referência: (RIBEIRO et al., 2016)

promissora no desenvolvimento de biocatalisadores com maior eficiência através de modulação positiva de sua atividade enzimática.

Essa modulação positiva em questão foi discutida anteriormente na seção 2.3. Há algumas décadas, se conseguiu criar acoplamentos alostéricos entre domínios de proteínas biologicamente não relacionadas usando diferentes abordagens que, geralmente, são baseadas em estratégias *in vitro* de evolução da proteína (DUEBER et al., 2003; EDWARDS et al., 2008; GOODEY; BENKOVIC, 2008; GUNTAS et al., 2005). A maioria destes estudos tem o objetivo de criar "proteínas interruptores", nas quais as proteínas mostram estados ativos ou inativos em resposta à vinculação de uma molécula, por exemplo. Entretanto, no contexto do coquetel enzimático utilizado na produção do etanol lignocelulósico, é crucial que as enzimas fiquem ativas durante todo o processo para melhorar sua eficiência.

Em face dessas duas questões, ou seja, a inibição da xilanase pelo seu próprio produto reacional e a possibilidade de criação de enzimas multidomínio interligadas para criação de uma modulação positiva, Ribeiro e colaboradores apresentaram em 2015 enzimas fundidas mais eficientes que a xilanase original isolada (RIBEIRO et al., 2015). Através de uma estratégia semirracional de *design* proteico, promoveram a inserção de uma xilanase GH11 de *Bacillus subtilis* (XynA\_WT) em uma proteína receptora de xilose (XBP, do inglês, *xylose binding protein*) através de *linkers* específicos (aminoácidos de cadeia lateral curta) que apresentaram uma estimulação alostérica em sua atividade catalítica causada pela xilose produzida.

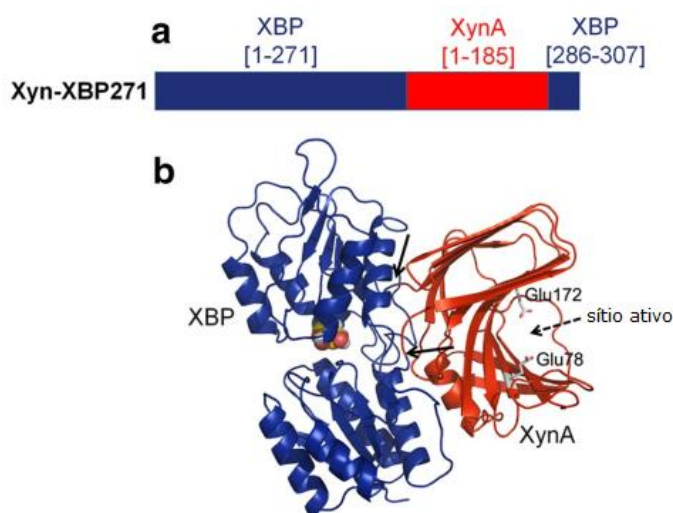
No ano seguinte, outro artigo do mesmo grupo foi publicado com estratégia similar, porém variando a posição na qual foi feita a inserção da XynA na estrutura da XBP (RIBEIRO et al., 2016). No primeiro artigo, duas posições de ligação haviam sido avaliadas, a 209 e 262, e neste último, a posição 271. A Fig. 6.1 ilustra o posicionamento dessas inserções. Para que ocorram essas melhorias na catálise enzimática, as simulações de dinâmica molecular (DM) podem, novamente, ser utilizadas, seja através do entendimento molecular da movimentação, conformação e exposição dos resíduos catalíticos, ou pela análise energética das interações substrato-enzima.

Após apresentação dos detalhes computacionais empregados nas simulações DM (seção 6.1), o capítulo trará detalhes das quimeras XynA-XBP ligadas na

posição 271 da XBP (seção 6.2) e posteriormente, dados comparativos entre quimeras ligadas nas diferentes posições (209 e 262), com e sem xilose (seção 6.3).

### 6.1 Detalhes computacionais

A quimera com a xilanase na posição 271 da XBP (XynA-XBP271) foi mais estável do que a xilanase isolada livre (XynA\_WT) a 55 °C, e na presença de xilose, sua eficiência catalítica foi cerca de três vezes maior que a XynA\_WT (RIBEIRO et al., 2016). Com o intuito de entender a nível molecular quais fatores responsáveis por esses resultados experimentais, em particular a movimentação, interações não ligadas na interface proteína-proteína e o efeito alostérico ativado pela presença de xilose, várias simulações DM foram realizadas.



**Figura 6.1** - Representação da quimera XynA-XBP. a) A sequência da quimera XynA-XBP271. Regiões da XBP estão apresentadas em azul e as da XynA em vermelho. O número entre colchetes indica o número do aminoácido nas proteínas isoladas. b) Representação tridimensional do modelo final obtido via simulações DM da quimera XynA-XBP271. A estrutura da XBP ligada à xilose está apresentada em azul e a da xilanase (XynA) em vermelho. As setas indicam os pontos de inserção da xilanase na XBP. O sítio ativo e os resíduos catalíticos (Glu78 e Glu172) no domínio xilanolítico estão também em destaque.

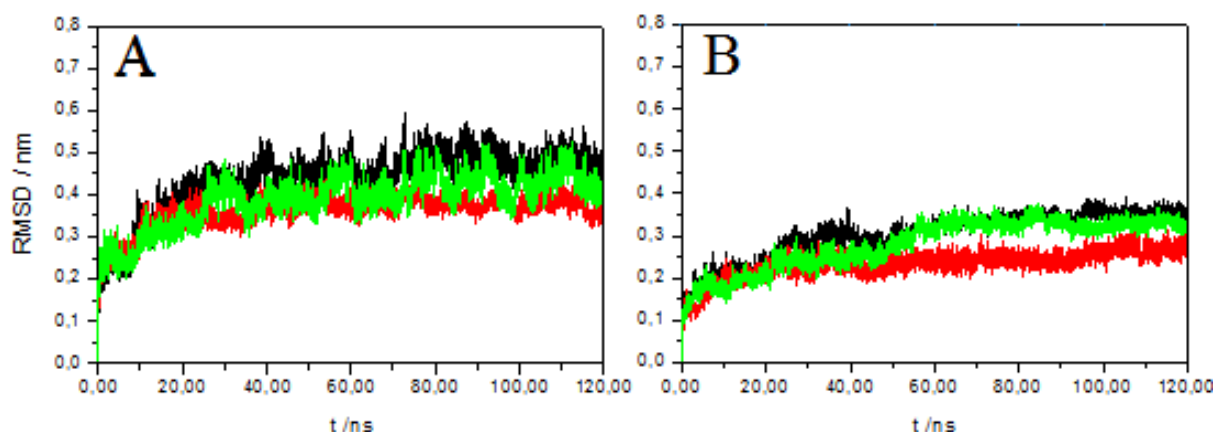
As estruturas das quimeras XynA-XBP271 com xilose (doravante XynA-XBP271\_cX) e sem xilose (doravante XynA-XBP271\_sX) no sítio ativo da XBP foram avaliadas usando o pacote de software GROMACS 5.0.2 (BERENDSEN; VAN DER SPOEL; VAN DRUNEN, 1995; PRONK et al., 2013; VAN DER SPOEL et al., 2005) e

o campo de força GROMOS-96, versão 53A6 (OOSTENBRINK et al., 2004). Ver seção 4.2 para mais detalhes. As coordenadas atômicas iniciais das quimeras foram obtidas através da modelagem por homologia empregando o programa MODELLER 9.13 (ŠALI et al., 1995) para mesclar o XynA (código pdb: 1XXN) e a XBP (códigos PDB: 3M9W, livre de xilose e 3MA0, ligado à xilose), respectivamente. Essas estruturas iniciais foram validadas usando o *software Procheck* (LASKOWSKI; MACARTHUR; THORNTON, 1998) com uma etapa subsequente de minimização de energia usando o método *steepest descent*. As estruturas resultantes foram solvatadas por moléculas de água (do tipo SPC, seção 4.1.2) em caixas de simulação dodecaédricas. Íons  $\text{Na}^+$  foram adicionados para garantir a eletroneutralidade dos sistemas. Dinâmicas de restrição de 400 ps foram realizadas a uma temperatura de referência de 300 K. Todos os sistemas foram simulados no *ensemble* NVT, pH neutro e 300 K, com um tempo total de 120 ns. A temperatura foi controlada pelo termostato V-rescale (BUSSI; DONADIO; PARRINELLO, 2007) e as ligações covalentes envolvendo átomos de hidrogênio na proteína e moléculas de água foram restringidas pelos algoritmos LINCS (HESS et al., 1997) e SETTLE (MIYAMOTO; KOLLMAN, 1992), respectivamente (seção 4.1.3). As equações de movimento foram resolvidas usando o método de integração Leap-Frog (VAN GUNSTEREN; BERENDSEN, 1988) e o tamanho do passo fixado em 2,0 fs (seção 4.1.1). As velocidades iniciais foram ajustadas usando a distribuição de Maxwell-Boltzmann. As interações de longo alcance foram modeladas usando o método PME (ver seção 4.2.6) com um corte de 1,2 nm (DARDEN; YORK; PEDERSEN, 1993). A energia potencial de interação (IPE) entre grupos também foi calculada como mostrado na seção 4.3.4. O *Alanine Scanning* computacional (seção 4.3.6) foi feito através do programa ROBETTA (KORTEMME; BAKER, 2002) para identificar resíduos mais importantes energeticamente na interface proteína-proteína das quimeras.

## 6.2 Quimeras XynA-XBP271\_cX e XynA-XBP271\_sX

### 6.2.1 Resultados e Discussão

A raiz do desvio quadrático médio (RMSD) foi calculada tomando como referência os carbonos alfa ao longo do tempo a fim de avaliarmos a estabilidade das estruturas geradas via simulações DM (Fig. 6.2). Esse critério estabeleceu um estágio de equilíbrio alcançado em torno de 50 ns para ambas as quimeras (com e sem xilose), com valores de RMSD entre 0,23 nm e 0,44 nm, respectivamente. Avaliados isoladamente, os domínios XBP e xilanase também se estabilizam por volta de 50 ns para XynA-XBP271\_cX e XynA-XBP271\_sX.

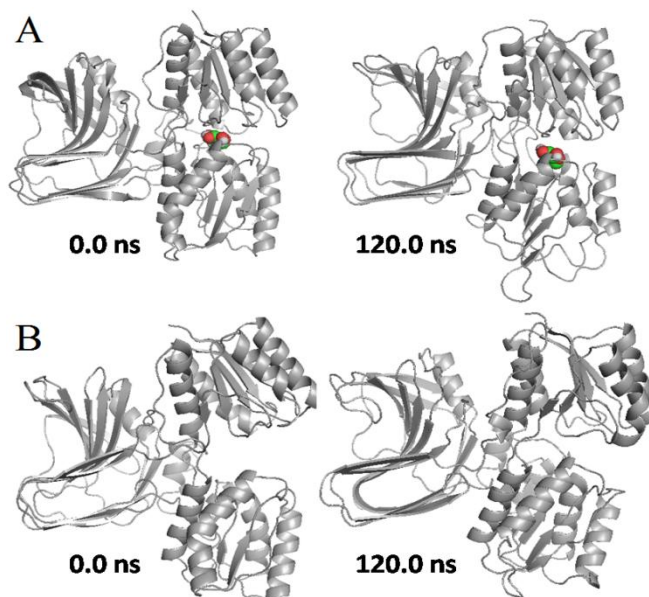


**Figura 6.2** - Raiz do desvio médio quadrático (RMSD) ao longo do tempo simulado da quimera XynA-XBP271 e dos seus domínios na a) ausência e b) presença de xilose no sítio de ligação da XBP. Linhas em preto indicam valores da quimera, em vermelho indicam valores do domínio xilanase e em verde referem-se ao domínio XBP.

Os valores de RMSD para a quimera ligada à xilose e seus domínios individuais são menores do que os dos sistemas sem xilose. Na quimera XynA-XBP271\_sX, a XBP apresenta uma conformação aberta e podem flutuar mais do que a mesma na forma fechada na qual existe a xilose ligada, logo, todas as estruturas e domínios, sofrem maiores variações. A Fig. 6.3 evidencia isso, apresentando as estruturas iniciais e finais para ambos estados da quimera XynA-XBP271.

As simulações DM sugerem a formação e estabilização da interface proteína-proteína entre os dois domínios da quimera XynA-XBP271, seja na ausência ou

presença da xilose no sítio de ligação da XBP, como pode ser visto na Fig. 6.3. Uma interface proteína-proteína similar também foi detectada em simulações DM em duas quimeras recentemente descritas por Ribeiro e colaboradores (RIBEIRO et al., 2015).

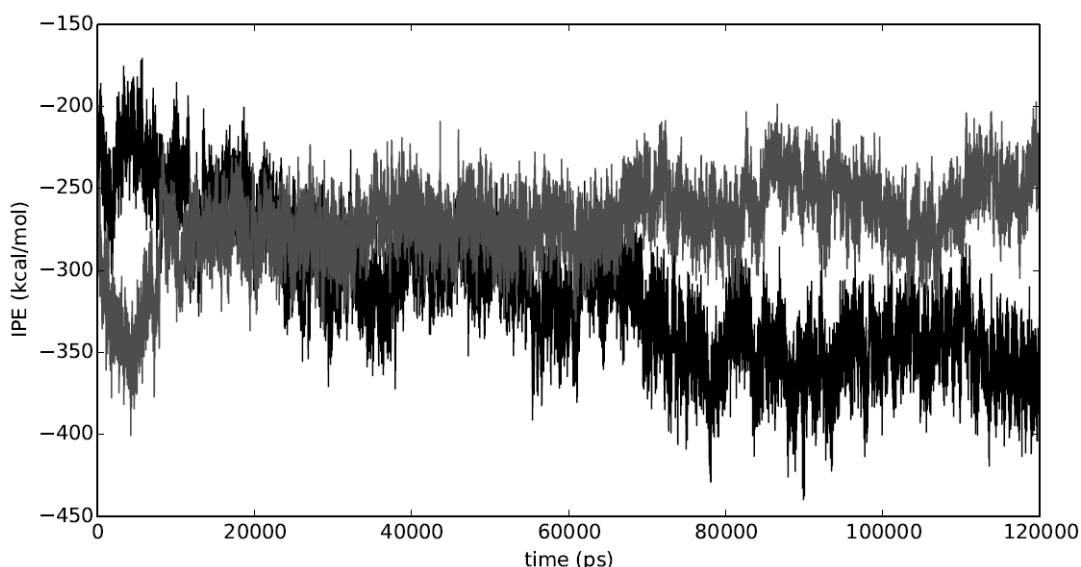


**Figura 6.3** - Estruturas iniciais e finais provenientes das simulações DM para a quimera XynA–XBP271 a) com xilose e b) sem xilose no sítio de ligação da XBP. Molécula de xilose representada em forma de bolas coloridas no sítio de ligação da XBP da imagem superior.

As energias potenciais de interação (IPE) das interfaces proteína-proteína interdomínios das quimeras XynA–XBP271 na presença e ausência de xilose foram calculadas e os resultados apresentados na Fig. 6.4. A variação da IPE ao longo do tempo simulado é um parâmetro confiável para análises de interface proteína-proteína (SHEINERMAN, 2000; SHEINERMAN; HONIG, 2002), e a média da IPE calculada para as quimeras (XynA–XBP271) com e sem xilose foi  $-311,78 \pm 43,74$  e  $-270,74 \pm 24,79$  kcal/mol, respectivamente. A comparação entre as energias de interação das interfaces da quimera XynA–XBP271 com as obtidas para as quimeras 2091A e 2621B (RIBEIRO et al., 2015), revelaram que a interface da quimera XynA–XBP271 ligada à xilose é aproximadamente 120 kcal/mol mais estável que a interface similar presente na quimera XynA-2621B ligada à xilose (RIBEIRO et al., 2015).

Portanto, a otimização da interface interdomínios como uma consequência da ligação da xilose é fortemente dependente da posição na qual a xilanase é fundida à XBP. Essa interface proteica na quimera XynA–XBP271 foi investigada em detalhes

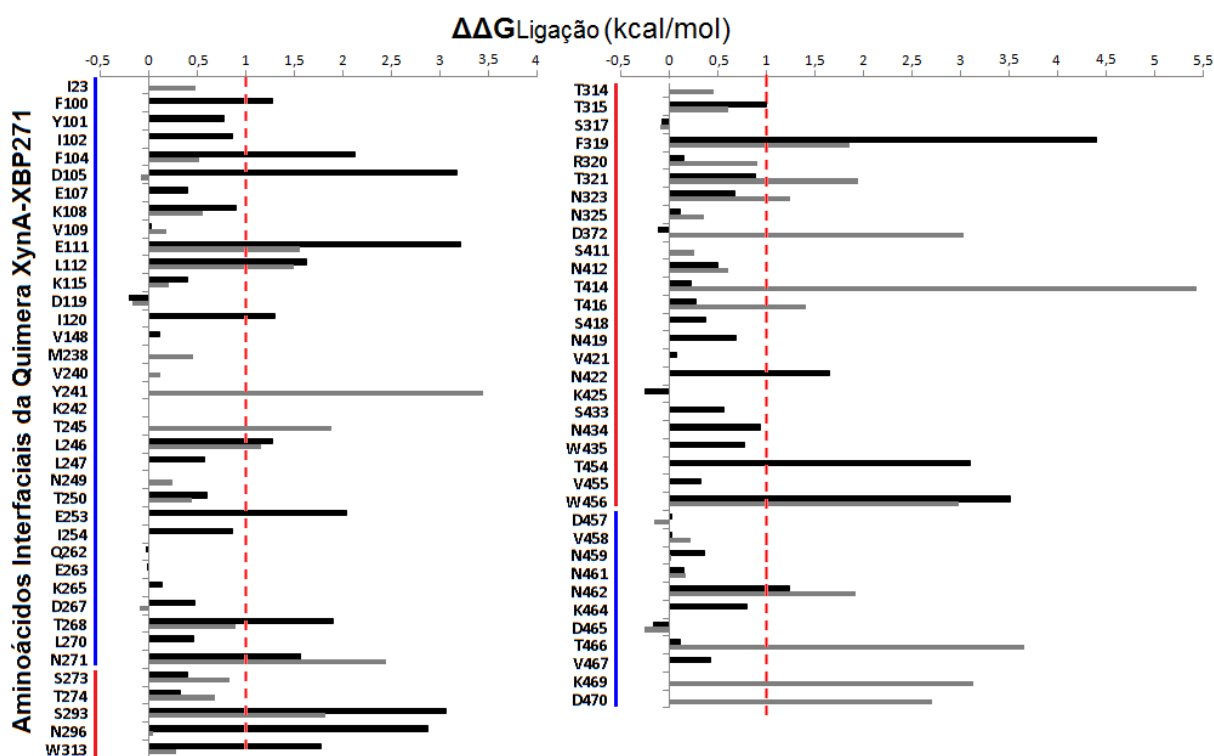
usando o *computational alanine scanning* (seção 4.3.6) para identificar os resíduos que mais contribuem para estabilização da interface interdomínios (Fig. 6.5).



**Figura 6.4** - Energia potencial de interação (IPE) interdomínios (xilanase e XBP) na quimera XynA–XBP271 com xilose (linha preta) e sem xilose (linha cinza). Dados obtidos via simulações DM ao longo de 120 ns.

Um aumento no  $\Delta\Delta G_{\text{ligação}}$  indica uma desestabilização da interface depois da substituição de um resíduo de alanina em uma determinada posição na qual se encontrava o aminoácido de interesse. Quando o valor do  $\Delta\Delta G_{\text{ligação}}$  pós-mutação excede 1,0 kcal/mol, a posição é considerada ser um *hot spot*, ou seja, um aminoácido muito importante para a estabilização da interface proteína-proteína.

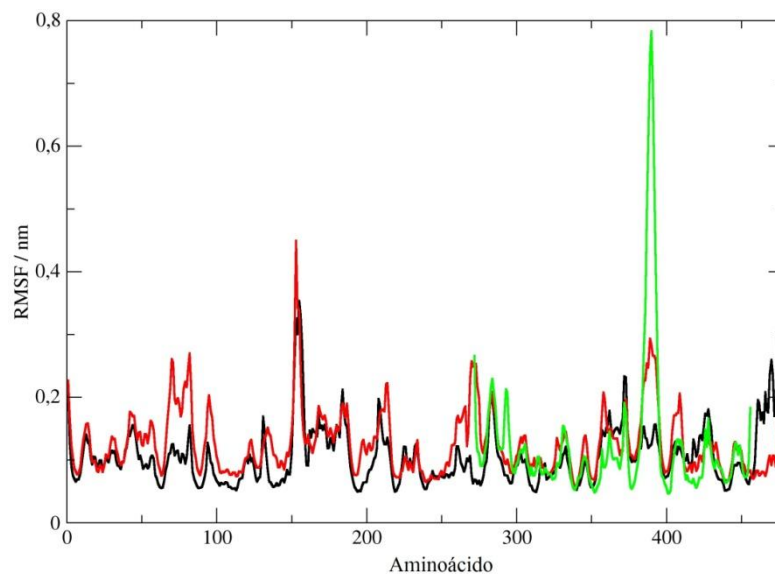
Para a quimera XynA–XBP271 na ausência de xilose, 41 resíduos foram identificados, sendo 18 deles considerados *hot spots*, com valores de  $\Delta\Delta G_{\text{ligação}}$  variando de 1,2 a 5,4 kcal/mol. Na mesma quimera com xilose ligada ao sítio da XBP, 65 resíduos foram identificados sendo 19 *hot spots* tendo valores de  $\Delta\Delta G_{\text{ligação}}$  variando de 1,0 a 4,4 kcal/mol. Os resíduos E111 e D105 (do domínio XBP), S293, F319, T454 e W456 (do domínio XynA) fazem uma contribuição significativa para a estabilidade da interface na quimera ligada a xilose, com valores de  $\Delta\Delta G_{\text{ligação}} > 3,0$  kcal/mol. Enquanto que para a quimera sem xilose ligada, os resíduos T466 (do domínio XBP) e o T414 (do domínio XynA) são os responsáveis por valores mais negativos com  $\Delta\Delta G_{\text{ligação}} > 3,5$  kcal/mol. Esses resíduos podem ser considerados os principais *hot spots* na interface interdomínios, e possuem, portanto, um papel fundamental para estabilização da interface proteína-proteína.



**Figura 6.5** - Resultado do *Alanine Scanning* computacional mostrando os resíduos mais importantes (acima do limiar de 1,0 kcal/mol indicado pela linha tracejada vermelha) para estabilização da interface proteína-proteína entre os domínios da XBP (destaque em azul) e Xilanase (destaque em vermelho) obtidos após simulações DM.  $\Delta\Delta G_{\text{ligação}}$  para mutações com alanina em resíduos da quimera XynA-XBP271 na presença (barras pretas) e ausência (barras cinza) de xilose.

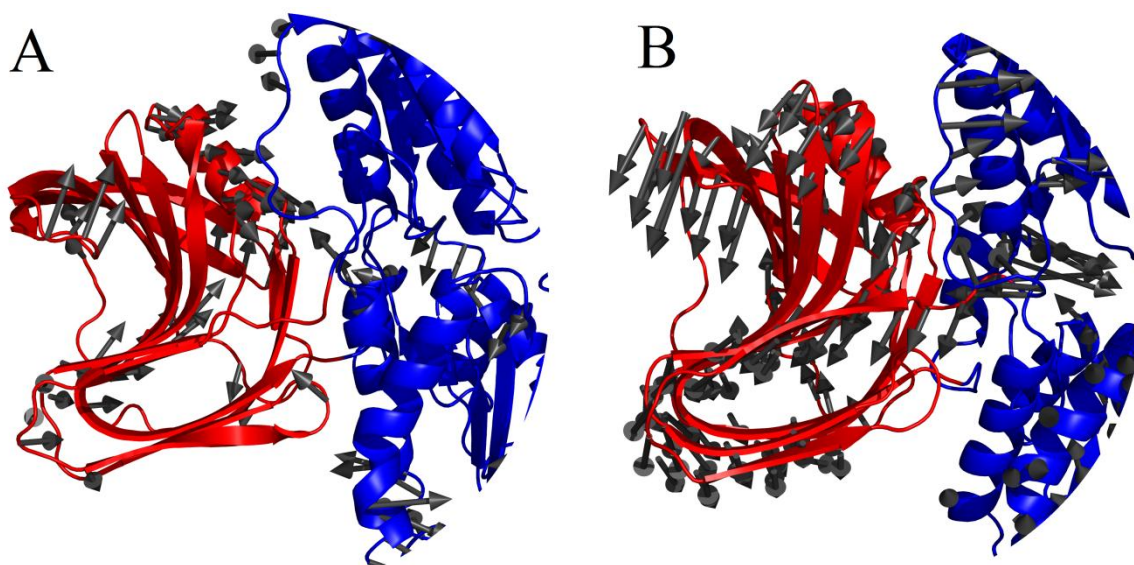
A flexibilidade da estrutura proteica foi quantitativamente analisada através da flutuação média quadrática (RMSF) por resíduo, utilizando o carbono alfa de cada aminoácido como referência para os cálculos. O RMSF calculado (Fig. 6.6) revela diferenças na flexibilidade nas posições do carbono alfa quando comparado a XynA\_WT e as quimeras XynA-XBP271 com e sem xilose. Variações significativas são observadas na região do polegar da xilanase (seção 3.2, resíduos 380 a 395) em todas as estruturas, sendo mais elevada na XynA\_WT, inferior na quimera XynA-XBP271\_sX e de menor valor na mesma quimera XynA-XBP271\_cX. Esses resultados indicam que a fusão da XBP à xilanase reduz significativamente a flexibilidade na região do polegar e que essa região torna-se ainda mais rígida com a presença da xilose.





**Figura 6.6** - Raiz da flutuação quadrática média (RMSF) por resíduo da quimera XynA-XBP271\_cX (linha preta), XynA-XBP271\_sX (linha vermelha) e XynA\_WT (linha verde).

Análises de *essential dynamics* da quimera XynA–XBP271 podem ser obtidas a partir das trajetórias das simulações DM e assim fornecer mais dados sobre as diferenças estruturais e funcionais observadas entre os dois estados das quimeras XynA–XBP271. Os principais movimentos ocorridos durante a simulação podem ser visualizados usando uma estrutura representativa proveniente das simulações DM na quais setas indicam a direção dos movimentos atômicos para destacar a dinâmica da proteína em larga escala em função da escala de tempo da simulação (Fig. 6.7). Movimentos diferentes no domínio xilanolítico podem ser observados na enzima com e sem xilose no sítio da XBP. Dados anteriores mostram uma transição entre as conformações "aberta" e "fechada" regulada por movimentos opostos da região do polegar e dos dedos (seção 3.2) para esse domínio (MURAKAMI et al., 2005; VIEIRA; DEGRÈVE; WARD, 2009). No entanto, esse movimento característico não está presente no domínio xilanase na enzima quimérica, onde foi observado um movimento de torção do domínio do polegar em relação às  $\beta$ -folhas do domínio dos dedos. Movimentos bem definidos do polegar foram observados tanto na presença como na ausência de xilose, no entanto, na ausência de xilose a fenda catalítica do domínio xilanolítico tornou-se mais exposto ao solvente.



**Figura 6.7** - *Essential dynamics* das quimeras representadas por setas em uma estrutura 3D representativa a partir das simulações DM. As setas indicam a direção dos movimentos locais das cadeias proteicas no domínio de xilanase (vermelha) nas quimeras a) XynA-XBP271\_cX e b) XynA-XBP271\_sX.

O volume do sítio ativo da xilanase foi calculado usando a ferramenta EPOSPB (EYRISCH; HELMS, 2007) e o resultado médio obtido para a quimera XynA-XBP271 foi um volume de 1150 Å<sup>3</sup> e 860 Å<sup>3</sup> para a XynA\_WT, demonstrando que o volume catalítico do sítio ativo é cerca de 25% maior na xilanase fundida do que a mesma na forma isolada. O resultado sugere que uma maior atividade xilanolítica observada na quimera pode ser resultado do aumento do acesso ao sítio ativo proporcionado ao substrato e é consistente com a redução nos valores experimentais obtidos nas enzimas quimerizadas (RIBEIRO et al., 2015).

Em geral, os resultados de simulação DM sugerem que a quimera XynA-XBP271 apresenta uma estrutura mais rígida em comparação com XynA\_WT e, em particular, a região do polegar demonstra uma redução significativa na flexibilidade. É sabido que essa região é a responsável pelo movimento abre-fecha do sítio catalítico, logo, sua exposição é diretamente relacionada com a atividade enzimática. A presença da xilose no domínio XBP proporcionou uma redução adicional na flexibilidade, não apenas no domínio da xilanase, mas em toda a estrutura da quimera. Embora a ligação da xilose reduzisse o movimento do domínio XBP, limitando sua estrutura à conformação fechada que é intrinsecamente menos flexível, os movimentos correlacionados no domínio da xilanase resultaram em um

aumento no volume da fenda catalítica, indicando que o acesso do substrato ao sítio ativo pode corresponder ao aumento da atividade xilanolítica observada na enzima quimérica. O efeito alostérico na quimera parece resultar de mudanças estruturais correlacionadas entre o domínio XBP e o domínio xilanolítico, e os dados da simulação DM identificam vários resíduos chave na interface proteína-proteína entre esses domínios que podem estar envolvidos na comunicação entre eles.

### 6.3 Comparações entre diferentes quimeras

Como exposto anteriormente, Ribeiro e colaboradores desenvolveram outras duas quimeras após a fusão da xilanase com XBP, diferindo somente a posição onde foi feita a inserção da xilanase na XBP. As três quimeras, portanto, possuem características diferentes das encontradas quando somente a xilanase livre era investigada. A fim de encontrar explicações para os resultados experimentais, e, a partir disso, propor melhorias para uma maior estabilidade e uma atividade aumentada, outras análises por DM foram realizadas. Para tal, a mesma metodologia empregada no artigo (RIBEIRO et al., 2015) foi replicada para as simulações das quimeras [209]XBP-AG-Xyn-GGA-XBP e [262]XBP-GGGG-Xyn-GA-XBP, doravante, XynA-XBP209 e XynA-XBP262, respectivamente. Para a simulação da xilanase isolada (XynA\_WT), os parâmetros de simulação utilizados foram os mesmos utilizados para a simulação da quimera XynA-XBP271, os quais constam na seção 6.1.

Na Tabela 6.1 a seguir, alguns dados experimentais de eficiência catalítica relativa e do efeito de ativação da xilose serão apresentados e nortearão as discussões a seguir, já que um dos objetivos das seções posteriores será a consolidação da explicação do melhor desempenho da xilanase após a fusão do domínio da XBP.

A eficiência catalítica (ECat), ou constante de especificidade, é a medida de quão eficiente uma enzima converte um substrato em produto. Neste caso, uma  $ECat_{relativa}$  será usada, ou seja, o valor se baseará na razão entre a eficiência catalítica da quimera (nos seus estados com ou sem xilose no sítio ativo da XBP) com o mesmo parâmetro da XynA\_WT. Assim podemos padronizar a análise que foi

feita com metodologias diferentes para os dois artigos publicados. A Eq. 1 detalha como o cálculo foi feito:

$$ECat_{relativa} = \frac{\left( \frac{k_{cat}}{K_M} \right)_{quimera}}{\left( \frac{k_{cat}}{K_M} \right)_{xilanase\ isolada}} \quad (1)$$

Para o efeito de ativação da xilose (EAX), o cálculo foi feito utilizando a Eq. 2, a partir da razão entre valores de eficiência catalítica relativa ( $ECat_{relativa}$ ) das quimeras com e sem xilose no sítio ativo da XBP. Esse dado pode revelar a influência da presença da xilose na  $ECat_{relativa}$  da quimera e tem fundamental importância para avaliação do efeito alostérico proposto.

$$EAX = \frac{(ECat_{relativa})_{com\ xilose}}{(ECat_{relativa})_{sem\ xilose}} \quad (2)$$

**Tabela 6.1** - Parâmetros cinéticos experimentais de eficiência catalítica ( $ECat_{relativa}$ ) e efeito de ativação da xilose para quimeras XynA-XBP209, XynA-XBP262 e XynA-XBP271.

Quimera	Eficiência Catalítica Relativa		Efeito de Ativação da Xilose	Referências
	sem xilose	com xilose		
XynA-XBP209	1,71	2,52	1,47	(RIBEIRO et al., 2015)
XynA-XBP262	1,29	2,27	1,76	
XynA-XBP271	1,83	3,31	1,81	(RIBEIRO et al., 2016)

\* Os valores de desvio-padrão foram omitidos, mas encontram-se nas respectivas referências.

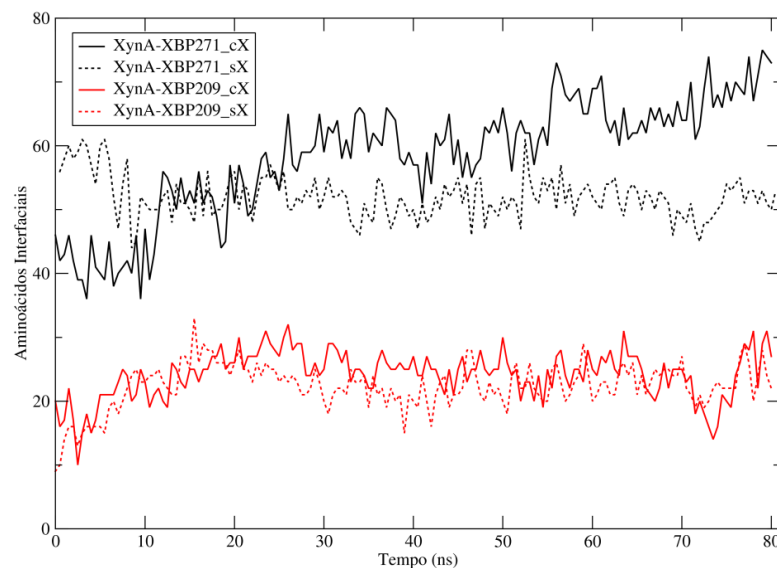
Os dados acima foram discutidos nos respectivos artigos publicados. Em resumo, mostram a maior eficiência catalítica da quimera XynA-XBP271 além de um efeito de ativação da xilose superior em relação às demais. Portanto, além de ser três vezes mais eficiente que a XynA\_WT, a XynA-XBP271 possui também um efeito alostérico propiciado pela entrada do próprio produto de catálise da xilanase. Isso a torna extremamente eficiente, mesmo quando comparada com quimeras muito eficazes, e possibilita sua utilização em coquetéis enzimáticos. As análises

comparativas entre as quimeras e a xilanase isolada serão tratadas nas seções a seguir para obtermos bases para entender como a mudança na posição da fusão do domínio xilanolítico na XBP pode produzir quimeras tão distintas. Nos artigos, a explicação para a diferença de atividade entre as quimeras e xilanase isolada baseou-se, dentre outros pontos, na menor flexibilidade encontrada na região do polegar (responsável pelo abre-fecha e consequente exposição do sítio ativo) e maior rigidez estrutural. A mesma análise estendeu-se para explicar o efeito alostérico causado pela entrada da xilose no sítio ativo da XBP. Outras análises serão apresentadas a seguir para corroborar com essa proposta já publicada.

### *6.3.1 Resultados e Discussão*

A quantidade de aminoácidos interfaciais foi calculada ao longo da simulação de acordo com a metodologia presente na seção 4.3.5. A diferença na quantidade de aminoácidos que interagem diretamente na interface proteína-proteína (xilanase-XPB) tem consequência direta do local no qual a fusão entre os domínios foi feita. É possível fazer uma correlação desse parâmetro com dois outros presentes na Tabela 6.1, a eficiência catalítica relativa e com o efeito de ativação causado pela xilose. A quantidade de aminoácidos interfaciais pode trazer indícios sobre possíveis mudanças estruturais na interface entre as enzimas após a entrada da xilose na XBP, e desse modo, fornecer dados para o entendimento do efeito alostérico.

A Fig. 6.8 traz a quantidade de aminoácidos interfaciais de duas quimeras, XynA-XBP209 e XynA-XBP271, com e sem xilose, identificados em 80 ns de simulação. As duas quimeras foram escolhidas por possuírem os valores de EAX mais discrepantes (~25%) e uma diferença na  $ECat_{relativa}$  de ~30% para as quimeras com xilose.



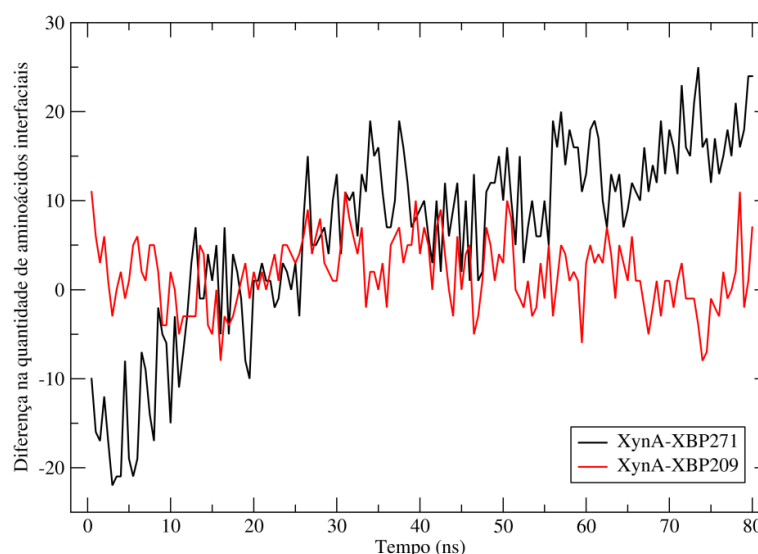
**Figura 6.8** - Quantidade de aminoácidos interfaciais nas quimeras XynA-XBP209\_cX (com xilose, vermelho contínuo), XynA-XBP209\_sX (sem xilose, vermelho pontilhado), XynA-XBP271\_cX (com xilose, preto contínuo) e XynA-XBP271\_sX (sem xilose, preto pontilhado). Dados coletados ao longo de 80 ns de acordo com metodologia presente na seção 4.3.5.

Os dados apresentados na Fig. 6.8 mostram uma quantidade superior de resíduos interfaciais para a quimera XynA-XBP271 em comparação à quimera XynA-XBP209 para ambos os casos, com e sem xilose. Uma estimativa da dimensão dessa interface pôde ser feita a partir da Fig. 6.4 nas quais dados de IPE mostraram uma maior estabilização entre domínios da quimera ligada à xilose em comparação da mesma sem a xilose no sítio de ligação da XBP. Entretanto, a IPE é uma medida indireta para extensão da interface já que a análise leva em consideração somente interações eletrostáticas que podem estar linearmente correlacionados com o tamanho da interface ou não. A partir da análise direta da quantidade de aminoácidos interfaciais esse problema é resolvido.

A maior extensão da interface na quimera XynA-XBP271, com cerca de 35 aminoácidos interfaciais a mais que a quimera XynA-XBP209, corrobora com as explicações fornecidas nas publicações e pode ser relacionada com a maior rigidez do domínio xilanolítico e consequente maior atividade. Com uma maior ancoragem proporcionada pela forte interação entre os domínios XBP e xilanase, a XynA-XBP271 apresenta uma menor flexibilidade assim como uma movimentação abre-fecha menos acentuada. Dessa forma, os resíduos catalíticos (glutamatos) ficam mais expostos ao longo do tempo na XynA-XBP271, favorecendo a reação.

A partir da Fig. 6.8, podemos gerar dados para correlacionar com os valores de EAX presentes na Tabela 6.1. A Fig. 6.9, construída a partir da figura anterior, traz a diferença entre a quantidade de aminoácidos interfaciais numa mesma quimera com e sem xilose ao longo do tempo simulado.

A diferença na quantidade de aminoácidos interfaciais ao longo do tempo de simulação revela a sensibilidade estrutural à presença de xilose no sítio ativo da XBP. A Fig. 6.2 apresenta uma constância em relação à quimera XynA-XBP209, a qual praticamente mantém a extensão da interface proteína-proteína na simulação com e sem xilose, com diferença entre os dois estados sempre próximo ao zero. Já a quimera XynA-XBP271 apresenta uma resposta positiva, com uma maior extensão da interface após a adição de xilose no sítio ativo da XBP, com cerca de 15 a 20 aminoácidos a mais ao final do tempo simulado.



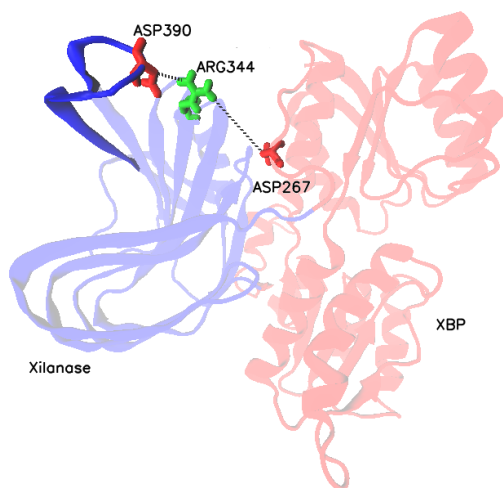
**Figura 6.9** - Diferença na quantidade de aminoácidos interfaciais nas quimeras com e sem xilose. Quimera XynA-XBP209 (vermelho) e XynA-XBP271 (preto). Dados coletados ao longo de 80 ns de acordo com metodologia presente na seção 4.3.5.

A análise do tempo total nos traz a informação que houve uma modulação positiva, ou seja, que o efeito alostérico causado pela adição de xilose realmente ficou evidenciado na quimera XynA-XBP271. Esse dado corrobora com o dado experimental de apresenta um aumento de cerca ~80% na eficiência catalítica relativa. Esse dado é mais um resultado que relaciona a extensão da interface com a atividade xilanolítica experimental.

As pontes salinas são interações eletrostáticas fortes que ocorrem na estrutura de proteínas entre dois aminoácidos os quais possuem cargas opostas, por exemplo, as interações entre as argininas e lisinas com aspartatos e glutamatos. Esse tipo de interação tem um poço energético mais profundo que as demais interações como ligações de hidrogênio ou interações de van der Waals e por isso, possuem papel de destaque tanto na estabilização estrutural quanto nas importantes movimentações funcionais.

Na quimera XynA-XBP271 foram identificadas algumas pontes salinas intradomínio e interdomínio, após 50 ns de simulação, que podem fornecer indícios do efeito alostérico positivo causado na atividade xilanolítica na presença da xilose.

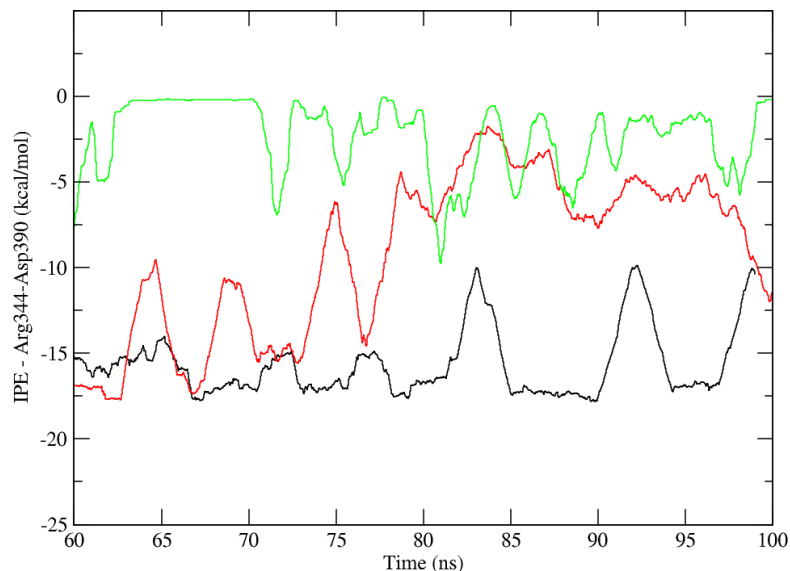
Análises de diversas pontes salinas encontradas ao longo da simulação DM da quimera XynA-XBP271, com e sem xilose, foram feitas com o intuito de correlacionar as suas funcionalidades com o EAX experimental. Foram encontradas duas interações desse tipo que indicavam uma dependência direta com a exposição do sítio ativo, logo, com a  $ECat_{relativa}$ . As pontes salinas, evidenciadas na Fig. 6.10, entre o Asp390-Arg344 e Asp267-Arg344, atuam diretamente na movimentação do polegar. Essa região já demonstrou ter uma alta flexibilidade, como apresentado na Fig. 6.6, com uma ordem crescente da quimera XynA-XBP271 com xilose, XynA-XBP271 sem xilose e a XynA\_WT.



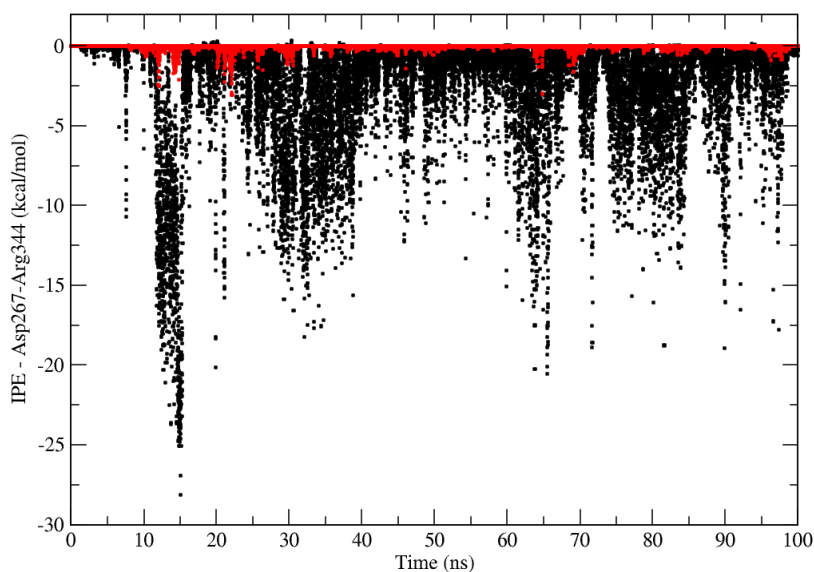
**Figura 6.10** - Representação de pontes salinas inter e intradomínios na quimera XynA-XBP271 com o domínio xilanolítico representado em azul (com a região do polegar destacada) e o domínio XBP, em vermelho. Aminoácidos Asp267 e Asp390 representados em forma de tubos vermelhos. Aminoácido Arg344 em forma de tubos verdes.



A energia potencial de interação dessas pontes salinas foi calculada em função do tempo de simulação para as quimeras XynA-XBP271 com e sem xilose. O resultado para a interação intradomínio Asp390-Arg344 foi mostrado a partir dos 60 ns (Fig. 6.11) após estabilização do RMSD de todos os sistemas (Fig. 6.2, para as quimeras XynA-XBP271).



**Figura 6.11** - Energia potencial de interação da ponte salina Asp390-Arg344 (em kcal/mol) na xilanase isolada (linha verde) e nas quimeras XynA-XBP271 com e sem xilose, XynA-XBP271\_cX (linha preta) e XynA-XBP271\_sX (linha vermelha), respectivamente. Dados coletados nos últimos 40 ns de simulação pós-estabilização do RMSD. Médias móveis a cada 500 frames.



**Figura 6.12** - Energia potencial de interação da ponte salina Asp267-Arg344 (em kcal/mol) nas quimeras XynA-XBP271 com e sem xilose, XynA-XBP271\_cX (linha preta) e XynA-XBP271\_sX (linha vermelha), respectivamente.

Além disso, um dado comparativo entre o mesmo parâmetro espacial foi feito para a XynA\_WT no caso da ponte salina intradomínio Asp390-Arg344. Na Fig. 6.12, resultados do mesmo parâmetro são apresentados para a ponte salina Asp267-Arg344. Nesse caso, não é possível avaliar o mesmo resultado para XynA-WT já que não existe um segundo domínio na xilanase isolada.

Os resultados referentes à ponte salina intradomínio corroboram com o dado de alta flexibilidade da região do polegar para XynA\_WT. Em vários momentos ao longo da simulação, a IPE entre esses dois aminoácidos para a xilanase livre chega à zero (ou próximo disso) indicando a ausência dessa interação. A oscilação percebida no gráfico condiz com o movimento abre-fecha do sítio catalítico característico das xilanases (VIEIRA; DEGRÈVE; WARD, 2009). Para a XynA-WT, além da magnitude da interação oscilar muito mais que o mesmo dado para as quimeras, os valores de menor energia atingidos estão em média próximos de -7 kcal/mol. Isso explica a alta movimentação do polegar que não consegue "ancorar-se" nessa ponte salina a qual possibilitaria uma abertura elevada do sítio, com consequente maior exposição dos resíduos catalíticos. Para a enzima quimerizada os valores de IPE são acentuadamente menores, logo, mais estáveis energeticamente, e as oscilações apresentam uma frequência menor. Apesar da quimera XynA-XBP271\_sX apresentar mais variações ao longo do tempo, ambas possuem valores menores para energia de interação dessa ponte salina. Para quimera com xilose, por exemplo, esses valores oscilam em média entre -17 a -10 kcal/mol indicando a permanência dessa ponte salina mesmo com pequenas movimentações, logo, o polegar se mantém numa posição de abertura do sítio ativo ao longo do tempo simulado.

A queda na flexibilidade na região do polegar que controla a exposição do sítio ativo, apresentada anteriormente em dados de RMSF (Fig. 6.6) e a movimentação de abertura do sítio através da movimentação do polegar, apresentada na Fig. 6.7 da análise da *essential dynamics* corroboram com o dado de energia potencial de interação da ponte salina interdomínios (Fig. 6.12). O mesmo resíduo de Arg344 do domínio xilanase interage com um Asp267 presente no domínio XBP de forma a manter essa abertura do sítio ativo, principalmente na quimera com xilose. A diferença na magnitude de interação, praticamente inexistente para a quimera sem xilose, favorece a rigidez do polegar, tornando a região

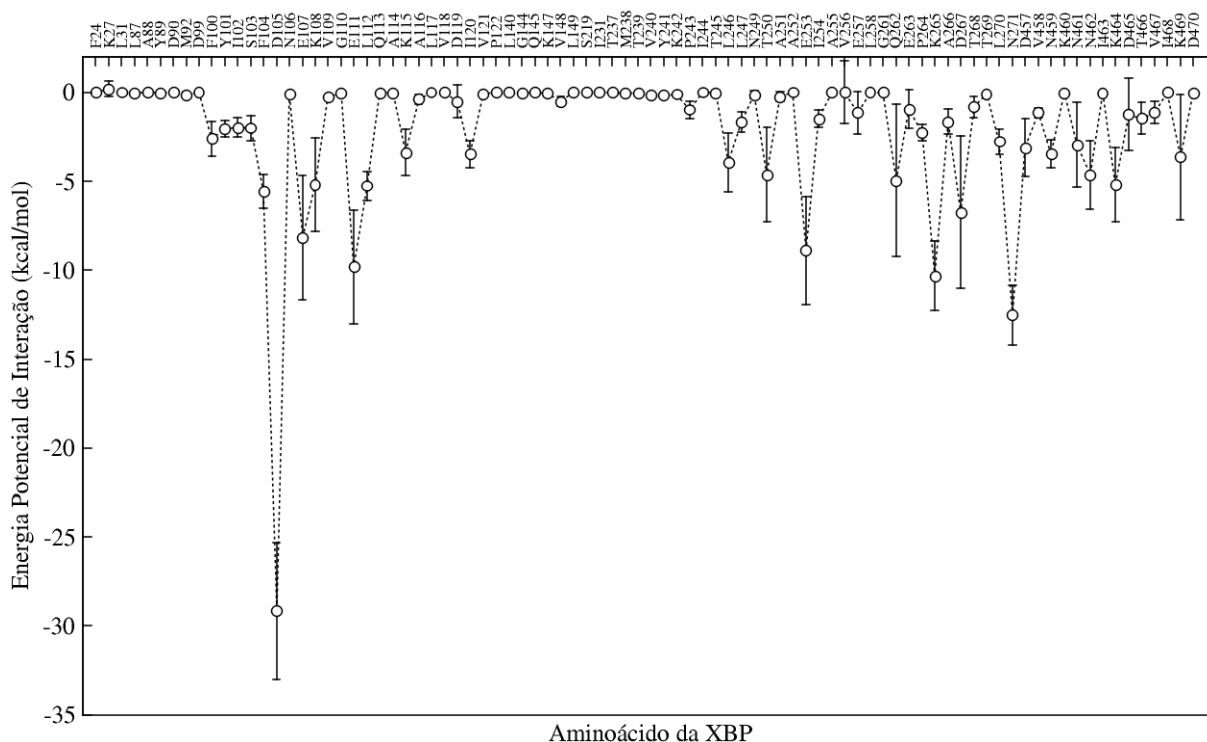
praticamente nula em movimentação. Apesar de ter valores médios inferiores quando comparados à interação intradomínio, a ponte salina Asp267-Arg344 é um fator importante para descrição do efeito alostérico já que a mesma interação praticamente não acontece na ausência de xilose no sítio ativo da XBP. Esses dados podem ser correlacionados com o valor de EAX da Tabela 6.1, que traz cerca de 80% de diferença entre os dois estados da quimera XynA-XBP271.

As interações entre resíduos na interface proteína-proteína são úteis na identificação de quais aminoácidos realmente desempenham a função estabilizante ou desestabilizante desses domínios. Resíduos que apresentam baixas energias de ligação com a vizinhança são identificados como importantes para que haja manutenção da interface, enquanto que o oposto torna o sistema instável energeticamente. Pelo observado (RIBEIRO et al., 2015, 2016), indicados na Tabela 6.1 e nas Fig. 6.8 e 6.9, o tamanho da interface é fator de correlação direta entre os dados moleculares de simulação DM e a eficiência catalítica relativa experimentalmente obtida. Logo, é importante a identificação da rede de interação entre os aminoácidos localizados nessa fronteira entre os domínios.

Uma análise com esse objetivo foi feita através do *alanine scanning* computacional, melhor detalhada na seção 4.3.6. Os resultados foram apresentados na seção 6.2.1 e Fig. 6.5 para as quimeras XynA-XBP271\_cX e XynA-XBP271\_sX. Entretanto, os valores possuem certa limitação já que os dados são obtidos a partir de uma única estrutura representativa das milhares resultantes da trajetória da simulação DM. Nesse intuito, cálculos de IPE foram feitos considerando a interação de todos os resíduos da xilanase presentes dentro de uma esfera de 8 Å a partir de cada resíduo da XBP nos últimos 50 ns de simulação, mostrados na Fig. 6.13. O contrário também foi feito, ou seja, todos os resíduos dentro da mesma distância de corte a partir da xilanase. Os resultados se mostraram muito similares e não foram inseridos nesse material.

Os resultados evidenciam a importância de determinados resíduos para estabilização da interface na quimera XynA-XBP271. Aproximadamente 50% desses resíduos não influenciam positivamente para isso, com valores próximos à zero para ambos os métodos de cálculo. Os outros 50% apresentam menores valores de IPE e são importantes para a manutenção da estrutura ligada. Em especial, os aminoácidos D105, E107, E111, E253, K265, D267, N271 e K464. Esses

aminoácidos, exceto a asparagina na 271, são do tipo carregados, no caso aspartatos, glutamatos e lisinas. Dessa forma, são propícios a interagir com aminoácidos também carregados ou polares no domínio xilanolítico. Logo, são fundamentais na estabilização da interface e consequentemente, no efeito alostérico positivo causado pela entrada da xilose no sítio ativo da XBP.



**Figura 6.13** - Energia potencial de interação entre os aminoácidos da interface que se distanciam no máximo 8 Å do outro domínio. Dados obtidos via simulação DM para os resíduos de XBP.

Análises como essa podem ser precursoras de estudos envolvendo mutações, como mostrado no capítulo anterior, já que fornecem indicativos de quais resíduos podem conferir maior estabilidade à interface, além de poder testá-los usando a mesma ferramenta sem custos adicionais. Com os dados obtidos então, encaminhar ao experimentalista para confirmar as hipóteses levantadas.

## 6.4 Conclusão

Os resultados obtidos via simulação DM para as quimeras XynA-XBP271 abrem possibilidades para a criação de enzimas multidomínios envolvendo a enzima xilanase e a XBP de forma racional. Novos análogos podem ser sugeridos com propriedades otimizadas em relação às atuais e assim aumentar a probabilidade de

uso em larga escala, principalmente pela análise da importância de determinado aminoácido na estabilização da estrutura formada.

A comparação entre as quimeras evidencia a correlação entre extensão da interface proteína-proteína e a atividade catalítica experimental. Além disso, os dados obtidos, em especial da interação entre aminoácidos de cargas opostas (ponte salina), nos permite entender o funcionamento dessas quimeras e a influência de um substrato para ativação do efeito alostérico.

## Referências - Capítulo 6

- ANDRIĆ, P. et al. Reactor design for minimizing product inhibition during enzymatic lignocellulose hydrolysis. **Biotechnology Advances**, v. 28, n. 3, p. 407–425, maio 2010.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1–3, p. 43–56, set. 1995.
- BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of chemical physics**, v. 126, n. 1, p. 014101, 7 jan. 2007.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089, 1993.
- DUEBER, J. E. et al. Reprogramming control of an allosteric signaling switch through modular recombination. **Science (New York, N.Y.)**, v. 301, n. 5641, p. 1904–8, 26 set. 2003.
- EDWARDS, W. R. et al. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. **Nucleic acids research**, v. 36, n. 13, p. e78, ago. 2008.
- EYRISCH, S.; HELMS, V. Transient Pockets on Protein Surfaces Involved in Protein–Protein Interaction. **Journal of Medicinal Chemistry**, v. 50, n. 15, p. 3457–3464, jul. 2007.
- GOODEY, N. M.; BENKOVIC, S. J. Allosteric regulation and catalysis emerge via a common route. **Nature chemical biology**, v. 4, n. 8, p. 474–82, ago. 2008.
- GUNTAS, G. et al. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 32, p. 11224–9, 9 ago. 2005.
- HESS, B. et al. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, set. 1997.
- JÖNSSON, L. J.; ALRIKSSON, B.; NILVEBRANT, N.-O. Bioconversion of lignocellulose: inhibitors and detoxification. **Biotechnology for Biofuels**, v. 6, n. 1, p. 16, 2013.
- KORTEMME, T.; BAKER, D. A simple physical model for binding energy hot spots in

protein-protein complexes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 22, p. 14116–14121, 2002.

LASKOWSKI, R. A.; MACARTHUR, M. W.; THORNTON, J. M. **Validation of protein models derived from experiment****Current Opinion in Structural Biology**, 1998.

MIYAMOTO, S.; KOLLMAN, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. **Journal of Computational Chemistry**, v. 13, n. 8, p. 952–962, out. 1992.

MURAKAMI, M. T. et al. Correlation of temperature induced conformation change with optimum catalytic activity in the recombinant G/11 xylanase A from *Bacillus subtilis* strain 168 (1A1). **FEBS Letters**, v. 579, n. 28, p. 6505–6510, 21 nov. 2005.

OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1656–1676, out. 2004.

PRONK, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics**, v. 29, n. 7, p. 845–854, 1 abr. 2013.

RIBEIRO, L. F. et al. Insertion of a xylanase in xylose binding protein results in a xylose-stimulated xylanase. **Biotechnology for Biofuels**, 2015.

RIBEIRO, L. F. et al. A xylose-stimulated xylanase-xylose binding protein chimera created by random nonhomologous recombination. **Biotechnology for biofuels**, v. 9, p. 119, 2016.

ŠALI, A. et al. Evaluation of comparative protein modeling by MODELLER. **Proteins: Structure, Function, and Genetics**, v. 23, n. 3, p. 318–326, nov. 1995.

SHEINERMAN, F. Electrostatic aspects of protein–protein interactions. **Current Opinion in Structural Biology**, v. 10, n. 2, p. 153–159, 1 abr. 2000.

SHEINERMAN, F. B.; HONIG, B. On the Role of Electrostatic Interactions in the Design of Protein–Protein Interfaces. **Journal of Molecular Biology**, v. 318, n. 1, p. 161–177, abr. 2002.

TEUGJAS, H.; VÄLJAMÄE, P. Product inhibition of cellulases studied with <sup>14</sup>C-labeled cellulose substrates. **Biotechnology for biofuels**, v. 6, n. 1, p. 104, 24 jul. 2013.

VAN DER SPOEL, D. et al. GROMACS: fast, flexible, and free. **Journal of computational chemistry**, v. 26, n. 16, p. 1701–18, dez. 2005.

VAN GUNSTEREN, W. F.; BERENDSEN, H. J. C. A Leap-frog Algorithm for

Stochastic Dynamics. **Molecular Simulation**, v. 1, n. 3, p. 173–185, mar. 1988.

VIEIRA, D. S.; DEGRÈVE, L.; WARD, R. J. Characterization of temperature dependent and substrate-binding cleft movements in *Bacillus circulans* family 11 xylanase: a molecular dynamics investigation. **Biochimica et biophysica acta**, v. 1790, n. 10, p. 1301–6, out. 2009.



---

## Capítulo 7

### **Dependência do pH na atividade xilanolítica de uma enzima multidomínio**

Enzimas diferentes podem ser fundidas de modo a criar uma estrutura multidomínio. Essa nova estrutura, também chamada de quimera, pode apresentar propriedades diferentes, em geral, melhores do que as originais. Alguns detalhes dessas enzimas multidomínios foram apresentados no Cap. 3 e uma grande aplicação biotecnológica foi apresentada e discutida no capítulo anterior (Cap.6). Trata-se da quimera (XynA-XBP271) formada após a fusão, na posição 271, de domínio xilanolítico (XynA) à uma proteína receptora de xilose (XBP) que apresentou uma eficiência catalítica experimental quase 3,5 vezes maior que a mesma xilanase não-quimerizada (Tabela 1, Cap. 6).

Dentre os vários dados experimentais apresentados na publicação de Ribeiro e colaboradores (2016), uma caracterização bioquímica foi feita para avaliar a atividade xilanolítica relativa da quimera XynA-XBP271 sob diferentes condições de temperatura, pH e tempo (RIBEIRO et al., 2016). Em especial, a análise de pH revelou uma grande deficiência da quimera em ambientes ácidos. Ambas as enzimas avaliadas nessa análise, XynA\_WT e XynA-XBP271 apresentaram pH ótimo em torno de 6 e 6,5, mas diminuem sua ação cerca de 80% e 50%, respectivamente, ao serem submetidas a pH 4,5, por exemplo (RIBEIRO et al., 2016). Apesar de resolver parcialmente o problema enfrentado pela XynA, a quimera ainda reage mal à condições ácidas.

Essa dependência à acidez do meio torna-se muito importante quando se pensa em qual ambiente elas são empregadas. Atualmente, os coquetéis

enzimáticos (tratados no Cap. 1 e no Cap.3) atuam no material lignocelulósico sólido que foi previamente tratado com agentes inorgânicos, em geral, ácidos fortes antes de serem submetidos à hidrólise enzimática. Devido essa alta dependência do pH apresentado pela maioria das enzimas presentes nos coquetéis, uma elevação do pH do sistema precisa ser feita para garantir o funcionamento adequado dessas enzimas. Esse procedimento acaba resultando em uma perda de tempo de produção e um maior custo associado ao processo. Portanto, é mais um entrave que têm sido combatido pelos usuários/defensores dessa metodologia, que como dito anteriormente, é mais ambientalmente correta quando comparada aos agentes inorgânicos nocivos.

Diante desse cenário, uma possibilidade seria criar enzimas (multidomínios ou não) com uma maior resistência à acidez do meio já que o processo sob baixos pHs já é consolidado e amplamente utilizado. Para tanto, é necessário que se entenda inicialmente as causas pelas quais essas enzimas perdem sua atividade tão vertiginosamente. Essas análises podem ser feitas empregando a simulação de dinâmica molecular (DM) sob diferentes condições.

### 7.1 Conceito das Simulações *Constant-pH MD* (CpHMD)

A consideração do pH em simulações DM de proteínas está relacionada diretamente aos valores de  $pK_a$  dos grupos ionizáveis das cadeias laterais. Esses dados são calculados experimentalmente e alocados nos campos de força para os cálculos dos parâmetros, já descritos no Cap. 4. Na forma convencional de cálculo, o estado de protonação das cadeias laterais é constante ao longo da simulação mesmo após as mudanças conformacionais observadas durante a movimentação de suas cadeias laterais, no caso de proteínas. Isso ocorre porque não há quebra/formação de ligações, já que todas as conectividades são referenciadas nos arquivos de campo de força preparados no início do cálculo e assim permanecem até o final da simulação.

Em face dessa questão, simulações conhecidas como *Constant-pH MD Simulations*, ou simplesmente, simulações CpHMD, foram desenvolvidas para o tratamento adequado dessas variações no pH, e, conseqüentemente, nos estados de protonação de grupos ionizáveis. Nos algoritmos de CpHMD, o estado de

protonação é atualizado durante a simulação, a fim de considerar o ambiente no qual determinado grupo ionizável se encontra no momento da atualização. Mais detalhes sobre essa abordagem nas simulações DM podem ser obtidas nas referências a seguir, como em *reviews* sobre o tema (CHEN et al., 2014; MONGAN; CASE, 2005; RADAK et al., 2017) e em publicações que utilizaram o método para estudos de enzimas e seu comportamento em diferentes pHs (BU et al., 2013; FUZO; DEGRÈVE, 2013; GERONIMO; PAYNE; SANDGREN, 2018; GRAMANY et al., 2016).

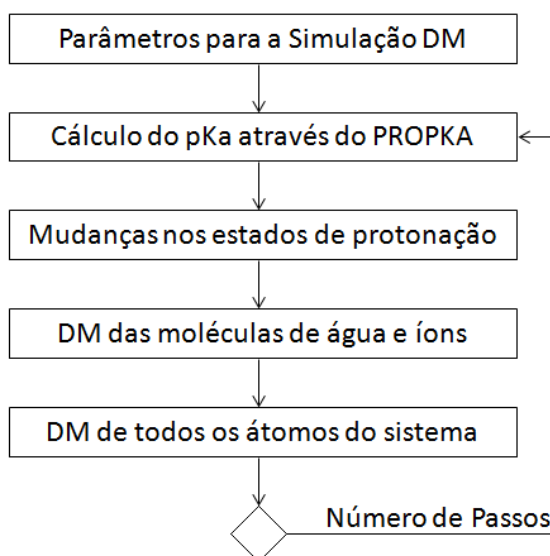
Vários métodos de CpHMD já foram publicados e a forma como o solvente é tratado difere bastante nos resultados obtidos. Vários métodos utilizando solvente implícito foram publicados (BARROSO DASILVA; DIAS, 2017; LEE; SALSBURY; BROOKS, 2004; MONGAN; CASE; MCCAMMON, 2004). Essa forma de tratamento do solvente foi abordada anteriormente no Cap. 4. Entretanto, métodos com solvente explícito trazem a possibilidade real do tratamento eletrostático das interações proteína-solvente, fundamentais para manutenção da estrutura proteica tridimensional. Essa abordagem também possui várias publicações como apresentadas nos artigos de revisão acima e nas referências a seguir, mais recentes (HUANG et al., 2016; SWAILS; YORK; ROITBERG, 2014).

## 7.2 Metodologia e parâmetros de simulação

A metodologia empregada para a simulação desta tese considera o solvente explicitamente, porém, devido ao alto custo computacional associado ao tratamento de milhares de moléculas de água, uma simplificação foi feita para que o cálculo se tornasse viável. O processo consiste em curtos ciclos de simulação DM divididos por intervalos nos quais a configuração dos estados de protonação são avaliados através do método empregado no programa PROPKA3 (OLSSON et al., 2011; SØNDERGAARD et al., 2011). A Fig. 7.1 descreve seu funcionamento (FUZO; DEGRÈVE, 2013).

Inicialmente os parâmetros para simulação são definidos e então o cálculo do pKa é feito através do PROPKA. De acordo com o pH desejado, ocorre a protonação nas cadeias laterais caso a condição de  $\text{pH} \leq \text{pKa}$  seja encontrada. Posteriormente uma simulação curta, somente com as moléculas de água e os íons livres para

movimentarem-se, é feita com duração de 0,2 ps e 1 fs como tamanho do passo. A intenção é que haja uma estabilização com os possíveis novos estados de protonação presentes nas cadeias laterais e as moléculas do solvente circundantes. Então, todo o sistema é liberado para movimentar-se numa simulação DM de 20 ps com tamanho do passo de 2 fs. O ciclo, portanto, é fechado e novamente a determinação dos estados de protonação é feita por N vezes.



**Figura 7.1** - Esquema representativo do funcionamento da simulação CpHMD utilizada nesta tese para determinar os estados de protonação das cadeias laterais na quimera XynA-XBP271 e da XynA\_WT. Imagem adaptada de (FUZO; DEGRÈVE, 2013).

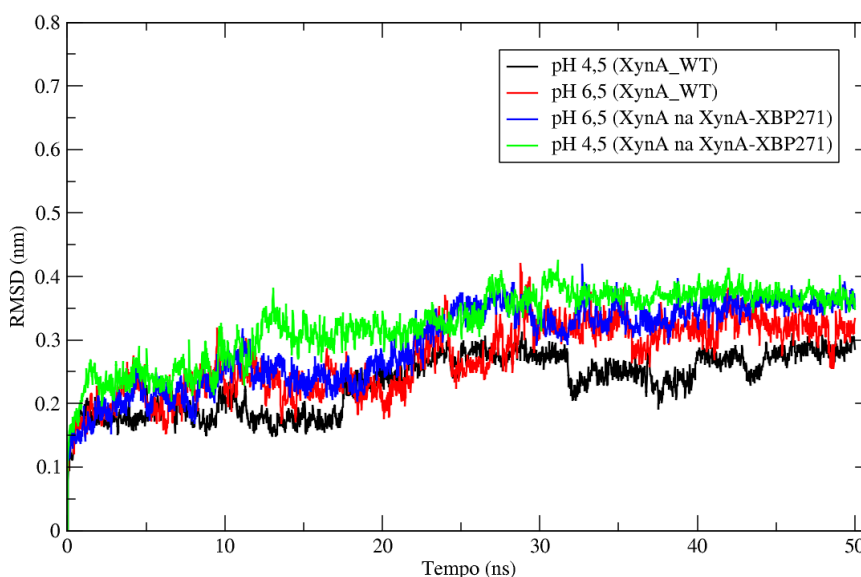
Através dessa metodologia, o tratamento da quimera XynA-XBP271 e da xilanase isolada (XynA\_WT) foi feita para o pH ótimo experimental (6,5) e para o pH 4,5, onde suas atividades apresentam acentuadas reduções. A estrutura de partida já havia passado por simulações DM todas descritas no capítulo anterior, com mais de 100 ns de simulação em pH neutro.

Utilizou-se o programa GROMACS e o campo de força GROMOS96 em 2500 ciclos (como os apresentados na Fig. 7.1) totalizando 50 ns de simulação total. O esquema de tratamento de cargas utilizado foi o Reaction Field já que em experiências prévias, o PME gerava artefatos que colapsavam a simulação DM. Um *cut-off* de 0,9 e 1,2 nm foram utilizados para limitar as interações de Coulomb e van der Waals, respectivamente. O modelo de água utilizado foi o SPC e a temperatura foi controlada através do termostato de Berendsen a 318 K. As ligações de

hidrogênio foram restringidas utilizando o algoritmo LINCS e o algoritmo SETTLE foi empregado para manter a estrutura das moléculas de água. A movimentação de todos os átomos foi integrada através do algoritmo *Leap-Frog* no *ensemble* NVT. Todos esses termos foram discutidos anteriormente no Cap. 4.

### 7.3 Resultados e Discussão

A análise de RMSD foi feita para os 50 ns simulados no intuito de avaliar a estabilidade das quimeras e da xilanase isolada ao longo da simulação e verificar a necessidade de estender a simulação a fim de que as estruturas se estabilizassem. A estrutura de partida utilizada foi a estrutura final obtida da simulação descrita no Cap. 6 e os carbonos alfa de cada aminoácido foram escolhidos como referência. A Fig. 7.2 apresenta os dados de RMSD para a quimera XynA-XBP271 e XynA\_WT em pH 4,5 e 6,5.

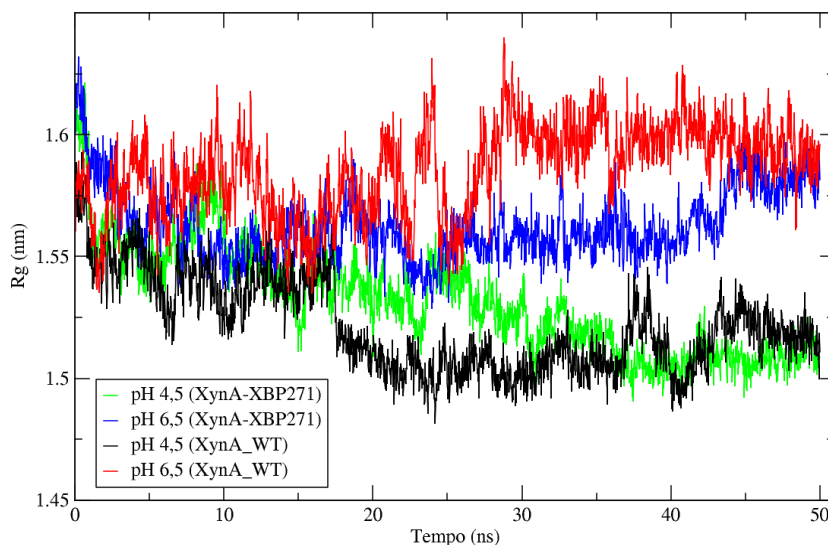


**Figura 7.2** - RMSD calculado tomando como referência os C $\alpha$  da quimera XynA-XBP271 em pH 4,5 (verde) e 6,5 (azul) e para XynA\_WT em pH 4,5 (preto) e 6,5 (vermelho) ao longo de 50 ns de simulação.

A Fig. 7.2 mostra que, após ~35 ns de simulação, as estruturas alcançaram um equilíbrio com RMSD próximo à 0,35 nm para a quimera em ambos os pH's, 0,32 nm para a XynA\_WT em pH 6,5 e 0,28 nm para o pH 4,5. A observação dos valores mais altos para a quimera é razoável já que a estrutura é maior devido ao domínio

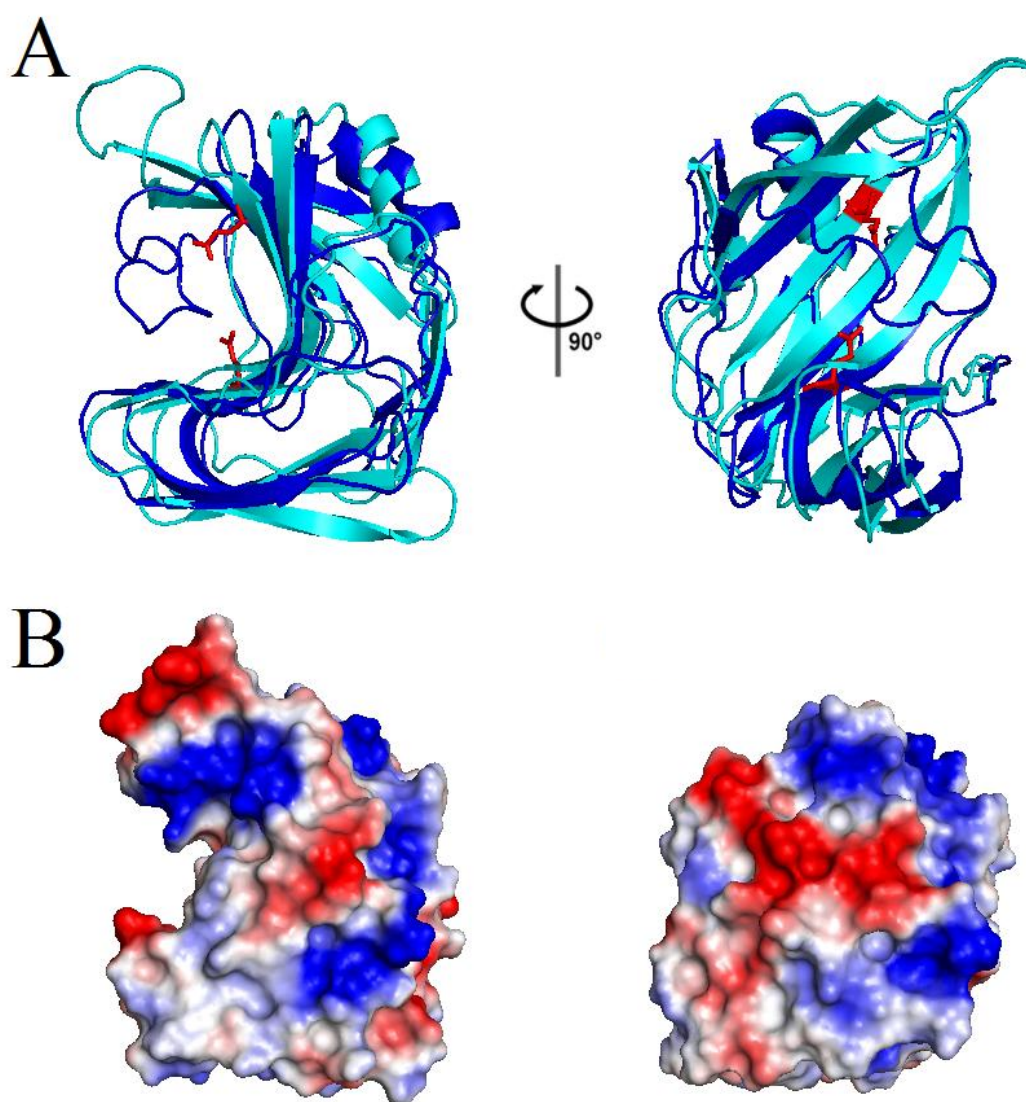
XBP e portanto é passível de maiores variações quando comparado ao sistema da XynA\_WT que possui somente um domínio. Além disso, apesar do tempo simulado não ser tão extenso como outros ao longo dessa tese, o resultado indica que os sistemas alcançaram estabilidade estrutural mesmo em um menor tempo de simulação e atesta a viabilidade de usar somente 50 ns nas análises a seguir.

O raio de giro pode ser usado como um indicativo da compactação da estrutura tridimensional de proteínas (LOBANOV; BOGATYREVA; GALZITSKAYA, 2008). Logo, altos valores de raio de giro indicam uma estrutura mais aberta, com cadeias mais distantes do centro de massa. Enquanto que valores menores indicam uma estrutura oposta, ou seja, com cadeias mais próximas de um ponto central e com maior grau de compactação. As xilanases GH11 apresentam uma estrutura similar a uma mão direita como apresentada no Cap. 3 . Seu sítio catalítico, formado por dois resíduos de glutamatos, estão próximos ao centro dessa estrutura, na superfície da região da palma. Logo, a análise do raio de giro pode indicar, mesmo que indiretamente, a exposição desse sítio visto que menores valores dessa medida apontam para uma estrutura mais fechada em torno desses aminoácidos. Do contrário, valores mais elevados sugerem uma estrutura mais aberta e disponível ao substrato. Essa medida já foi empregada para diversas enzimas estudadas através de simulações DM, inclusive xilanases (GRAMANY et al., 2016; KHAN et al., 2016; VENTORIM et al., 2018). A Fig. 7.3 apresenta o raio de giro ( $R_g$ ) em função do tempo simulado para as quatro situações já descritas acima.



**Figura 7.3** - Raio de giro calculado em longo de 50 ns para a quimera XynA-XBP271 em pH 4,5 (verde) e 6,5 (azul) e para XynA\_WT em pH 4,5 (preto) e 6,5 (vermelho).

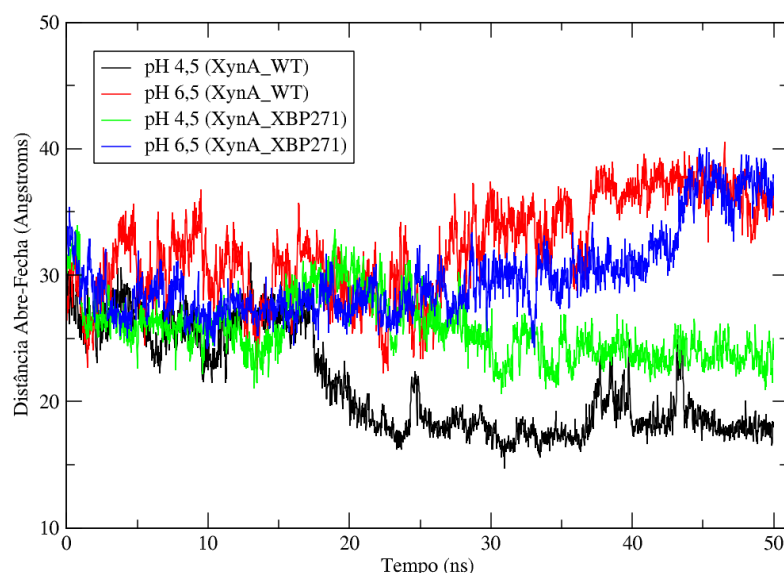
A análise indica que inicialmente os sistemas possuíam mesmo grau de compactação, contudo, pouco após o início da simulação os sistemas em pH mais ácido (4,5) tornaram-se mais compactos enquanto que os sistemas em pH 6,5 tornaram-se mais abertos, considerando apenas o domínio xilanolítico. Para ilustrar esse efeito da mudança do pH, a Fig. 7.4 mostra as estruturas finais das simulações para o caso da XynA\_WT nos diferentes pHs. A mudança de pH introduz hidrogênios em aminoácidos que antes estavam carregados ou neutros, no caso da histidina. Isso possibilita a interação através de ligações de hidrogênio tanto com moléculas do solvente quanto com outras cadeias laterais de aminoácidos acarretando em mudanças estruturais.



**Figura 7.4** - Representações da enzima XynA\_WT em diferentes pHs. A) Estruturas tridimensionais finais após 50 ns de simulação DM sobrepostas em dois ângulos distintos (visão lateral à esquerda e

visão frontal do sítio ativo à direita). Os dois resíduos catalíticos (E78 e E172) estão evidenciados em tubos vermelhos. Estrutura em pH 4,5 em azul claro e pH 6,5 em azul escuro. B) Superfície de potencial eletrostático para XynA\_WT em pH 6,5 (à esquerda) e pH 4,5 (à direita). Visão lateral do sítio ativo. Variação do vermelho (cargas mais negativas) para o azul (cargas mais positivas).

A Fig. 7.4 além de ilustrar a compactação da xilanase isolada em diferentes pHs, evidencia a diferença na exposição do sítio ativo das xilanases. A disponibilidade do sítio ativo dessas enzimas às moléculas do substrato pode ser avaliada através da distância abre-fecha do polegar à região dos dedos (VIEIRA; DEGRÈVE; WARD, 2009). A Fig. 7.5 traz esse resultado para a XynA\_WT e XynA\_XBP271 ao longo da simulação para os dois pHs discutidos nesse capítulo.



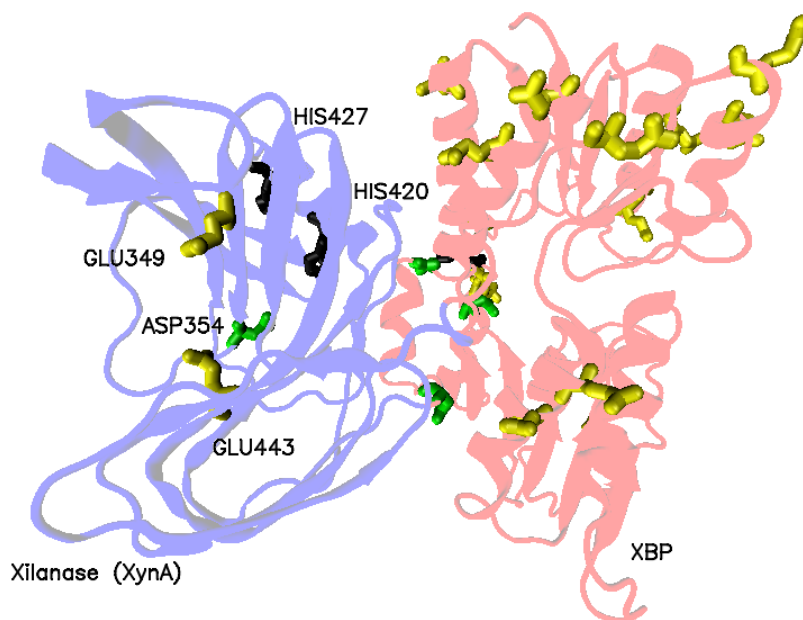
**Figura 7.5** - Distância Abre-Fecha (em Angstroms) em longo de 50 ns para a quimera XynA-XBP271 em pH 4,5 (verde) e 6,5 (azul) e para XynA\_WT em pH 4,5 (preto) e 6,5 (vermelho). A medida avaliada entre o polegar e a região dos dedos (Fig. 3.4).

A distância abre-fecha que descreve o movimento do polegar reitera os dados apresentados para o raio de giro. Ou seja, em pH 4,5, tanto a quimera como a xilanase isolada apresentam uma distância menor entre o polegar e os dedos, o que significa menor exposição do sítio catalítico. Enquanto que esses valores são superiores em pH 6,5 para XynA\_WT e XynA\_XBP271. Essa é a principal movimentação das xilanases e é uma característica dessa família de enzimas.

A fim de buscar explicações para diferença de atividade observada e dos parâmetros calculados teoricamente apresentados acima, os estados de protonação



de vários resíduos foram avaliados durante o tempo simulado. Diversos aminoácidos mudaram seu estado de protonação quando comparamos o pH 6,5 e o pH 4,5. A Fig. 7.6 (representando a estrutura XynA-XBP271) demonstra quais resíduos sofreram alterações no estado de protonação de sua cadeia lateral durante essa mudança e a Tabela 7.1 detalha quantitativamente essa variação ao apresentar a fração de tempo que o resíduo esteve protonado durante a simulação. A coluna da enzima isolada (XynA\_WT) só constará os aminoácidos do domínio XynA. Para a quimera XynA\_XBP271, os resíduos entre 1-271 e 457-478 pertencem ao domínio XBP enquanto que os resíduos entre 272-456 pertencem ao domínio xilanolítico, como foi apresentado na Fig. 6.1. Na simulação da XynA\_WT, que não possui o domínio XBP, os mesmos aminoácidos destacados no domínio XynA da Fig. 7.6 sofreram alteração durante a simulação. Contudo, possuem a numeração diferente devido a estrutura primária não ser igual. Os aminoácidos Asp83, Glu78, Glu172, His149 e His156 correspondem respectivamente à Asp354, Glu349, Glu443, His420 e His427 na imagem. Doravante, a numeração da quimera XynA\_XBP271 será utilizada a fim de padronizar a comparação entre os resíduos.



**Figura 7.6** - Representação dos aminoácidos que tiveram o estado de protonação das cadeias laterais alterados quando comparadas as simulações em pH 6,5 e 4,5 da quimera XynA\_XBP271. Em destaque os aminoácidos do domínio da xilanase (XynA, em azul). Histidinas representadas em tubos pretos, glutamatos em tubos amarelos e aspartatos em tubos verdes. Os aminoácidos Asp354, Glu349, Glu443, His420 e His427 correspondem respectivamente à Asp83, Glu78, Glu172, His149 e His156 na simulação da XynA\_WT.

A variação de pH 6,5 para 4,5 só afetou aminoácidos de histidina, aspartato e glutamato já que os pKas de suas cadeias laterais estão próximos à essa variação, no caso  $6,6 \pm 1,0$ ,  $3,5 \pm 1,2$  e  $4,2 \pm 0,9$  (GRIMSLEY; SCHOLTZ; PACE, 2009). Os aminoácidos destacados na Fig. 7.6 são pertencentes ao domínio xilanolítico, His420, His427, Asp354 e os dois aminoácidos catalíticos, Glu443 e Glu349 que atuam diretamente na reação de quebra da xilana em xilose.

**Tabela 7.1** - Fração de tempo que as cadeias laterais dos resíduos estiveram na sua forma protonada. Valores descritos para a quimera XynA\_XBP271 e XynA\_WT em pH 6,5 e 4,5.

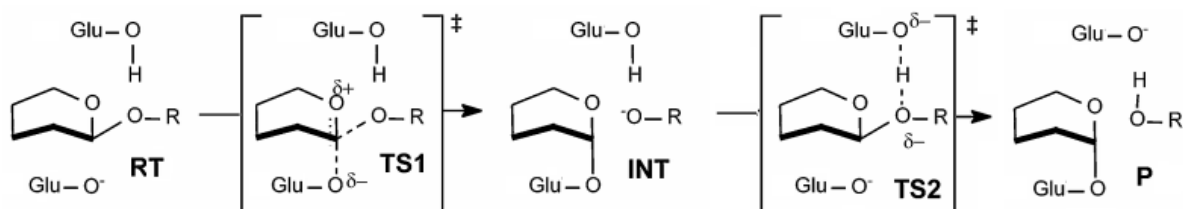
	XynA_XBP271		XynA_WT	XynA_XBP271		XynA_WT	
	pH 4,5	pH 6,5		pH 4,5	pH 6,5	pH 4,5	pH 6,5
<b>Glu15</b>	19,8%	0,0%	(a)	<b>Glu253</b>	83,4%	0,0%	
<b>Glu29</b>	98,2%	0,0%		<b>Glu257</b>	78,4%	0,0%	(a)
<b>Glu44</b>	32,5%	0,0%		<b>Glu263</b>	57,8%	0,0%	
<b>Glu45</b>	88,2%	0,0%		<b>Glu349</b>	100%	0,4%	100% 0,7%
<b>Glu52</b>	78,5%	0,0%		<b>Asp354</b>	70,1%	0,0%	86,8% 0,0%
<b>Glu78</b>	89,9%	0,0%		<b>His420</b>	79,7%	0,0%	36,4% 0,0%
<b>Glu82</b>	96,1%	0,0%		<b>His427</b>	100%	13,4%	100% 13,8%
<b>Glu107</b>	69,3%	0,0%		<b>Glu443</b>	100%	14,3%	100% 40,5%
<b>Glu111</b>	28,7%	0,0%		<b>Asp470</b>	74,1%	22,9%	
<b>Glu172</b>	20,2%	0,0%		<b>His473</b>	100%	35,2%	(a)
<b>Glu179</b>	83,3%	0,0%		<b>Glu475</b>	37,3%	0,0%	
<b>Asp224</b>	75,8%	0,0%		<b>Glu477</b>	43,7%	0,0%	

(a) Aminoácidos não existem nessa enzima pois pertencem ao domínio XBP

Na Tabela 7.1 podemos confirmar a capacidade preditiva da simulação CpHMD no tratamento dos estados de protonação das cadeias laterais dos aminoácidos descritos no parágrafo anterior. Dentre os 478 aminoácidos da quimera XynA-XBP271, 24 deles, presentes na tabela acima, alteraram seus estados de protonação. Para a XynA\_WT, 6 dos 185 aminoácidos também alteraram seus estados de protonação. Outros aminoácidos tiveram menos que 5% de variação do estado protonado para desprotonado (ou o contrário) não foram considerados na Tabela 7.1. Os aminoácidos no domínio XBP da quimera que apresentaram

modificações no estado de protonação das cadeias laterais encontram-se distantes do domínio XynA ou não apresentaram modificações importantes para a atividade xilanolítica observada.

É sabido que em xilanases da família GH11 como a XynA\_WT e XynA na forma de quimera, o mecanismo para quebra do substrato xilana ocorre por meio dos dois glutamatos agindo sinergicamente (Fig. 7.7). Um deles age como base de Lewis e outro como ácido de Brønsted-Lowry. Seus pKa's, medidos através de RMN  $^{13}\text{C}$ , foram 4,6 e 6,7 (MCINTOSH et al., 1996). Fazendo a correspondência da cadeia primária com os dados experimentais, pode-se garantir que para o caso do domínio xilanase na quimera XynA-XBP271 esses valores referem-se aos Glu349 e Glu443, e na XynA\_WT aos Glu78 e Glu172.



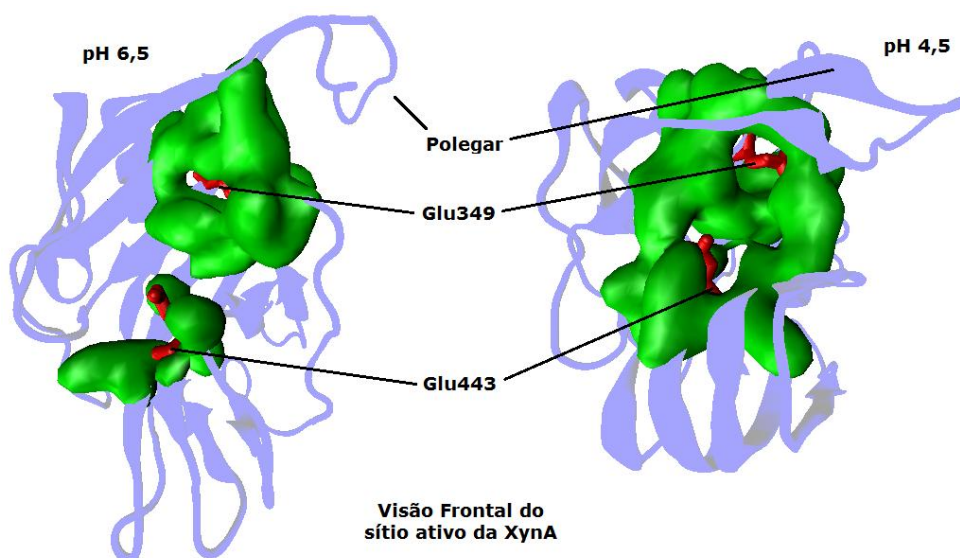
**Figura 7.7** - Mecanismo de reação representativo para formação do complexo enzima-substrato. Substrato xilana representada como grupo R. Etapas envolvendo reagentes (RT), estados de transição (TS1 e TS2), intermediário (INT) e produto (P) até esse ponto da reação.

Como mostrado na Fig. 7.7, para que a reação ocorra no pH ótimo, um dos glutamatos deve estar na sua forma desprotonada, atuando como nucleófilo, neste caso o Glu349, enquanto que o Glu443 deve estar protonado agindo como ácido de Brønsted. É exatamente isso que nos revela as simulações CpHMD realizadas. Ou seja, em pH 6,5 o Glu349 que possui pKa 4,6 está majoritariamente na forma desprotonada (0,4% e 0,7% para XynA\_XBP271 e XynA\_WT respectivamente) enquanto que o Glu443 (pKa 6,7) possui parte de suas moléculas na forma protonada (14,3% e 40,5% para XynA\_XBP271 e XynA\_WT respectivamente). Em pH mais ácido, ambos tornaram-se protonados nesse tempo simulado, inviabilizando o ataque nucleofílico que inicia a quebra do substrato.

A avaliação da área da superfície acessível ao solvente (ASAS, seção 4.3.7) apresenta uma informação importante acerca da atividade xilanolítica desempenhada basicamente pelos dois glutamatos catalíticos, Glu349 e Glu443, como já mencionado no Cap. 3 e ao longo desse documento. Essa propriedade

indica a área da superfície do aminoácido que pode ser acessada por moléculas do solvente. Uma correlação direta desse dado pode ser feita com a exposição do resíduo às moléculas presentes no solvente, consequentemente, o substrato. Um aminoácido que tenha baixa ASAS está blindado e não pode ser acessado devido ao enovelamento proteico ou qualquer outro impedimento espacial.

O aminoácido Glu349 apresenta ASAS de  $0,03 \pm 0,03 \text{ nm}^2$  em pH 6,5 e 4,5. Em contrapartida, o aminoácido Glu443 tem ASAS igual a  $0,40 \pm 0,20 \text{ nm}^2$  em pH 6,5 e  $0,03 \pm 0,03 \text{ nm}^2$  para o pH mais ácido. Logo, no pH ótimo, a quimera XynA-XBP271 apresenta um dos resíduos catalíticos cerca de 10 vezes mais exposto do que sob condições mais ácidas, como ilustrado na Fig. 7.8.



**Figura 7.8** - Representação da visão frontal do sítio ativo da XynA (da quimera XynA-XBP271) destacando a blindagem dos glutamatos catalíticos para os pHs 6,5 (à esquerda) e 4,5 (à direita). Estrutura secundária em azul, resíduos catalíticos, Glu349 e Glu443, em tubos vermelhos e uma superfície representativa da blindagem existente (em verde).

A Fig. 7.8 ilustra os resultados de ASAS com o resíduo Glu443 tendo maior possibilidade de interação com o solvente (ou substrato) em pH 6,5 e a mudança brusca no seu impedimento espacial para o pH mais ácido.

Além da exposição quase nula dos aminoácidos catalíticos em pH 4,5, a análise dos estados de protonação desses dois glutamatos também fornece indícios do motivo da diminuição da ação xilanolítica em um ambiente ácido.

Portanto, as simulações em CpHMD utilizando o método em questão forneceram dados compatíveis com os obtidos experimentalmente e se credenciam a futuras utilizações para simulações do mesmo sistema ou outros sistemas proteicos. A His420 por exemplo, teve seu estado de protonação alterado, saindo de totalmente desprotonada em pH 6,5 (em ambos os sistemas) pra 79,7% do tempo no estado protonado na quimera XynA\_XBP271 e menos da metade disso (36,4% do tempo no estado protonado) na xilanase isolada e pode ser um aminoácido importante para explicar a maior resistência das quimeras em relação à xilanase isolada que é observada experimentalmente, visto que ela se encontra muito próximo à interface entre os dois domínios e compõe a fita A-4 (Fig. 5.1), a mesma que pertence um dos glutamatos catalíticos (Glu443). Outro aminoácido que deve ser melhor investigado é o Asp354 que altera quase que totalmente seu estado de protonação e encontra-se na interface entre os domínios.

#### 7.4 Conclusão

Os dados obtidos podem ser utilizados como base para desenvolvimento de novas xilanases ou quimeras através da proposição de mutações a fim de aumentar a resistência à meios mais ácidos. O mecanismo experimental de quebra do substrato com os respectivos estados de protonação dos aminoácidos envolvidos foram replicados com êxito pelas simulações CpHMD. Comparações entre quimeras e domínios isolados podem ser feitos previamente visando a otimização de parâmetros relacionados à variação de pH tão importantes em mecanismos enzimáticos. Para tanto, mais análises precisam ser feitas para que seja definido um mecanismo pelo qual novas otimizações sejam desenvolvidas, sendo esses resultados somente os precursores e não os definitivos nesse sentido.

## Referências - Capítulo 7

- BARROSO DASILVA, F. L.; DIAS, L. G. Development of constant-pH simulation methods in implicit solvent and applications in biomolecular systems. **Biophysical reviews**, v. 9, n. 5, p. 699–728, out. 2017.
- BU, L. et al. Computational Investigation of the pH Dependence of Loop Flexibility and Catalytic Function in Glycoside Hydrolases. **Journal of Biological Chemistry**, v. 288, n. 17, p. 12175–12186, 26 abr. 2013.
- CHEN, W. et al. Recent development and application of constant pH molecular dynamics. **Molecular simulation**, v. 40, n. 10–11, p. 830–838, 1 jan. 2014.
- FUZO, C. A.; DEGRÈVE, L. The pH dependence of flavivirus envelope protein structure: insights from molecular dynamics simulations. **Journal of Biomolecular Structure and Dynamics**, v. 32, n. 10, p. 1563–1574, 3 out. 2013.
- GERONIMO, I.; PAYNE, C. M.; SANDGREN, M. The role of catalytic residue p K a on the hydrolysis/transglycosylation partition in family 3  $\beta$ -glucosidases. **Organic & Biomolecular Chemistry**, v. 16, n. 2, p. 316–324, 2018.
- GRAMANY, V. et al. Cloning, expression, and molecular dynamics simulations of a xylosidase obtained from *Thermomyces lanuginosus*. **Journal of Biomolecular Structure and Dynamics**, v. 34, n. 8, p. 1681–1692, 2 ago. 2016.
- GRIMSLEY, G. R.; SCHOLTZ, J. M.; PACE, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. **Protein science : a publication of the Protein Society**, v. 18, n. 1, p. 247–51, jan. 2009.
- HUANG, Y. et al. All-Atom Continuous Constant pH Molecular Dynamics With Particle Mesh Ewald and Titratable Water. **Journal of Chemical Theory and Computation**, v. 12, n. 11, p. 5411–5421, 8 nov. 2016.
- KHAN, F. I. et al. A pH based molecular dynamics simulations of chitinase II isolated from *Thermomyces lanuginosus* SSBP. **Cogent Biology**, v. 2, n. 1, 25 abr. 2016.
- LEE, M. S.; SALSBUURY, F. R.; BROOKS, C. L. Constant-pH molecular dynamics using continuous titration coordinates. **Proteins**, v. 56, n. 4, p. 738–52, 1 set. 2004.
- LOBANOV, M. Y.; BOGATYREVA, N. S.; GALZITSKAYA, O. V. Radius of gyration as an indicator of protein structure compactness. **Molecular Biology**, v. 42, n. 4, p. 623–628, 10 ago. 2008.
- MCINTOSH, L. P. et al. The pKa of the general acid/base carboxyl group of a

glycosidase cycles during catalysis: a  $^{13}\text{C}$ -NMR study of bacillus circulans xylanase. **Biochemistry**, v. 35, n. 31, p. 9958–66, 6 ago. 1996.

MONGAN, J.; CASE, D. A. Biomolecular simulations at constant pH. **Current opinion in structural biology**, v. 15, n. 2, p. 157–63, abr. 2005.

MONGAN, J.; CASE, D. A.; MCCAMMON, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. **Journal of computational chemistry**, v. 25, n. 16, p. 2038–48, dez. 2004.

OLSSON, M. H. M. et al. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. **Journal of Chemical Theory and Computation**, v. 7, n. 2, p. 525–537, 8 fev. 2011.

RADAK, B. K. et al. Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems. **Journal of Chemical Theory and Computation**, v. 13, n. 12, p. 5933–5944, 12 dez. 2017.

RIBEIRO, L. F. et al. A xylose-stimulated xylanase-xylose binding protein chimera created by random nonhomologous recombination. **Biotechnology for biofuels**, v. 9, p. 119, 2016.

SØNDERGAARD, C. R. et al. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. **Journal of Chemical Theory and Computation**, v. 7, n. 7, p. 2284–2295, 12 jul. 2011.

SWAILS, J. M.; YORK, D. M.; ROITBERG, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. **Journal of Chemical Theory and Computation**, v. 10, n. 3, p. 1341–1352, 11 mar. 2014.

VENTORIM, R. Z. et al. Impact of the removal of N-terminal non-structured amino acids on activity and stability of xylanases from Orpinomyces sp. PC-2. **International Journal of Biological Macromolecules**, v. 106, p. 312–319, jan. 2018.

VIEIRA, D. S.; DEGRÈVE, L.; WARD, R. J. Characterization of temperature dependent and substrate-binding cleft movements in Bacillus circulans family 11 xylanase: a molecular dynamics investigation. **Biochimica et biophysica acta**, v. 1790, n. 10, p. 1301–6, out. 2009.

---

## Conclusões Gerais

Nesta tese, alguns aspectos de interesse biotecnológico de enzimas xilanases da família GH11 foram estudados através de simulações computacionais, e, por meio dos resultados, foi possível explicar dados experimentais que somente poderiam ser entendidos se analisados a nível molecular. No Cap. 5, simulações DM em diferentes temperaturas catalíticas ótimas foram empregadas para entender o papel de mutações pontuais em aminoácidos que resultaram em um aumento da termoestabilidade de uma enzima xilanase em 20 °C. Dados de flexibilidade e interações intra e intermoleculares em determinadas regiões podem explicar a melhor interação dos aminoácidos modificados com outros resíduos próximos.

No Cap.6, o parâmetro estudado foi a atividade catalítica desempenhada pelo domínio xilanolítico. Quimeras criadas após a fusão de uma enzima xilanase GH11 e uma proteína receptora de xilose em diferentes posições foram comparadas para melhor entendimento dos dados experimentais que indicavam uma resposta catalítica até 3,5 vezes maior após a fusão dos domínios. Em consequência, o efeito alostérico causado pela entrada da xilose também pôde ser entendido por um aumento na rigidez estrutural que torna o sítio catalítico mais exposto quando comparado a xilanase livre. O tamanho da interface formada entre as duas enzimas também fornece indícios da correlação entre o tamanho da interface e a atividade catalítica relativa.

Por fim, o Cap. 7 trouxe dados teóricos sobre o comportamento da quimera XynA-XBP271 e XynA\_WT em diferentes pHs a fim de fundamentar quais aspectos envolvidos na queda da atividade dessas enzimas em ambientes ácidos. Uma menor exposição dos resíduos catalíticos mostrou ser crucial para que a atividade xilanolítica caia nos dados da quimera e a mudança no estado de protonação dos resíduos catalíticos, que dessa forma, impede a ação xilanolítica, indica o mecanismo da perda de função. Além disso, aminoácidos que modificaram as



cargas de suas cadeias laterais durante o tempo simulado dão indícios de serem alvos para futuras mutações a fim de desenvolver enzimas mais resistentes à queda no pH do meio.

Todos os resultados apresentados também atestam a capacidade descrita das simulações computacionais através de métodos clássicos ou usando *scripts* para avaliação de diferentes pHs. A partir dos resultados apresentados nesta tese, outros estudos de engenharia de proteínas, especialmente em xilanases, podem ser desenvolvidos a fim de otimizar ainda mais sua aplicação biotecnológica através de uma maior resistência a ambientes ácidos e de maior temperatura, apresentando alta eficiência catalítica.